THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 45, PART 2 March 1, 1933 No. 247

CONTENTS

	PAGE
LEWIS F. RICHARDSON. Time-marking a cathode-ray oscillogram	135
E. G. RICHARDSON and E. TYLER. The flow of liquid suspensions	142
H. L. WRIGHT. The influence of atmospheric suspensoids upon the earth's electric field as indicated by observations at Kew Observatory	152
P. VIGOUREUX and S. WATTS. The temperature coefficient of the saturated Weston cell	172
L. F. BATES. A new apparatus for the measurement of the earth's magnetic field	180
ALLAN FERGUSON and J. T. MILLER. A method for the determination of the specific heats of liquids, and a determination of the specific heats of aniline and benzene over the approximate range 20° C. to 50° C.	194
E. V. APPLETON and GEOFFREY BUILDER. The ionosphere as a doubly refracting medium	208
J. F. HERD. The generation and reception of wireless signals of short duration	221
R. NAISMITH. A comparison of the frequency-change and group-retardation methods of measuring ionized-layer equivalent heights	235
MARY TAYLOR. The Appleton-Hartree formula and dispersion curves for the propagation of electromagnetic waves through an ionized medium in the presence of an external magnetic field. Part 1: Curves for zero absorption	245
LEWIS SIMONS and E. H. SMART. A model to illustrate the motion of a diatomic rotator with two degrees of freedom	266
Sir A. S. EDDINGTON. Notes on the method of least squares	271
G. I. FINCH and R. W. SUTTON. The control of ignition-coil discharge characteristics	288
F. J. W. WHIPPLE. The wet-and-dry-bulb hygrometer: the relation to theory of the experimental researches of Awbery and Griffiths	307
MARY BELL and S. E. GREEN. On radiometer action and the pressure of radiation	320
Reviews of books	358

Price to non-Fellows 7/- net; post free 7/3

Annual subscription 35/- post free, payable in advance

Published by

THE PHYSICAL SOCIETY

1 Lowther Gardens, Exhibition Road

London, S.W.7

Printed at

THE UNIVERSITY PRESS, CAMBRIDGE

THE PHYSICAL SOCIETY

1932–33: OFFICERS OF THE SOCIETY

President:—Prof. A. O. RANKINE, O.B.E., D.Sc., F.Inst.P.

Hon. Secretaries:

EZER GRIFFITHS, D.Sc., F.Inst.P., F.R.S. (*Business*).

A. FERGUSON, M.A., D.Sc., F.Inst.P. (*Papers*), 70 *Hadham Road, Bishops Stortford, Herts.*

Hon. Foreign Secretary:—Prof. O. W. RICHARDSON, M.A., D.Sc., F.R.S.

Hon. Treasurer:—R. S. WHIPPLE, M.I.E.E., F.Inst.P.

Hon. Librarian:—J. H. BRINKWORTH, D.Sc., A.R.C.S.

Assistant Secretary:—H. R. LANG, PH.D., A.R.C.S., F.Inst.P.

Office of the Society:—1 *Lowther Gardens, Exhibition Road, London, S.W.* 7.

Editor:—Capt. C. W. HUME, M.C., B.Sc.,
14 *The Hawthorns, Finchley, N.* 3.

All communications, other than those to the Papers Secretary or the Editor, should be sent to the office of the Society, viz. 1 Lowther Gardens, Exhibition Road, London, S.W. 7.

INSTRUCTIONS TO AUTHORS

NOTE. The acceptance of a paper for publication in the Proceedings rests with the Council, advised by its Editing Committee. The high cost of printing renders it imperative to exclude matter that is not novel and not of importance to the understanding of the paper.

Authors offering original contributions for publication in the Proceedings should observe the following directions; failure to comply with these may cause considerable delay in publication. Fuller directions, including special instructions on the arrangement of mathematical work, are contained in a leaflet obtainable from the Assistant Secretary.

Manuscript.—A clear and concise style should be adopted, and the utmost brevity consistent with effective presentation of the original subject-matter should be used. The copy should be easily legible, preferably typewritten and double-spaced. It should receive a careful final revision before communication, since alterations are costly when once the type has been set up. Mathematical expressions should be set out clearly, in the simplest possible notation.

References.—In references to published papers the author's initials and name followed by the title of the journal in italics, volume, page and year should be given thus: *Proc. Phys. Soc.* 43, 199 (1931).

Drawings and tables.—Diagrams must be carefully drawn in Indian ink on white paper or card. Their size and thickness of line must be sufficient to allow of reduction. *Lettering and numbering should be in pencil,* to allow of printing in a uniform style. The number of diagrams should be kept down to the minimum. Data should in general be presented in the form of either curves or tables, but not both. Footlines descriptive of figures, and headlines indicative of contents of tables, should be supplied. *Sheets should not be larger than foolscap.*

Abstracts.—Every paper must be accompanied by an abstract in duplicate, brief but sufficient to indicate the scope of the paper and to summarize all novel results.

Proofs.—Proofs of accepted papers will be forwarded to authors. They should be returned promptly with errors corrected, but additions to or other deviations from the original copy should be avoided.

Reprints.—Fifty copies of printed papers will be supplied gratis. Extra copies may be purchased at cost price.

Contributions by non-Fellows.—Papers by non-Fellows must be communicated to the Society through a Fellow.

Republication.—Permission to reproduce papers or illustrations contained therein may be granted by the Council on application to the Hon. Secretaries.

THE INTERNATIONAL DECIMAL CLASSIFICATION OF EACH PAPER IS GIVEN ABOVE THE TITLE

MACMILLAN

INTRODUCTION TO THEORETICAL PHYSICS

(in five volumes)

By MAX PLANCK

Nobel Laureate, Foreign Member Royal Society, Professor of Theoretical Physics, University of Berlin, and President of the Kaiser Wilhelm Research Institute.

Translated by HENRY L. BROSE, M.A., D.Phil. (Oxon.), D.Sc.,
Lancashire-Spencer Professor of Physics, University College, Nottingham.

GENERAL MECHANICS
12s. net.

MECHANICS OF DEFORMABLE BODIES
10s. 6d. net.

THEORY OF ELECTRICITY AND MAGNETISM
10s. 6d. net.

THEORY OF LIGHT
10s. 6d. net.

THEORY OF HEAT
10s. 6d. net.

THE FIVE VOLUMES ARE AVAILABLE SEPARATELY

Prospectus post free on application.

A DICTIONARY OF APPLIED PHYSICS

EDITED BY SIR RICHARD GLAZEBROOK
K.C.B., D.Sc., F.R.S.

Vol. I. MECHANICS—ENGINEERING—HEAT

Vol. II. ELECTRICITY

Vol. III. METEOROLOGY, METROLOGY, AND MEASURING APPARATUS

Vol. IV. LIGHT, SOUND AND RADIOLOGY

Vol. V. AERONAUTICS, METALLURGY—GENERAL INDEX

In 5 Vols. £3. 3s. net each. The Set of 5 Vols. £14. 14s. net.

"In years to come the publication of this monumental work will rank as one of the milestones in British applied science." — NATURE

Descriptive Prospectus post free on application.

Macmillan & Co. Ltd. London W.C.2

DUBILIER

Metallized RESISTANCES

Noiseless

The background noise developed in these Resistances is extremely low.

Where there is a by-pass condenser, as in anode feed and grid bias circuits, the noise even after several stages of amplification, is too small to be measured.

Constant

When tested with the rated load applied for 16 hours each day, the resistance value *always* comes back to the initial value after cooling.

Non-Inductive

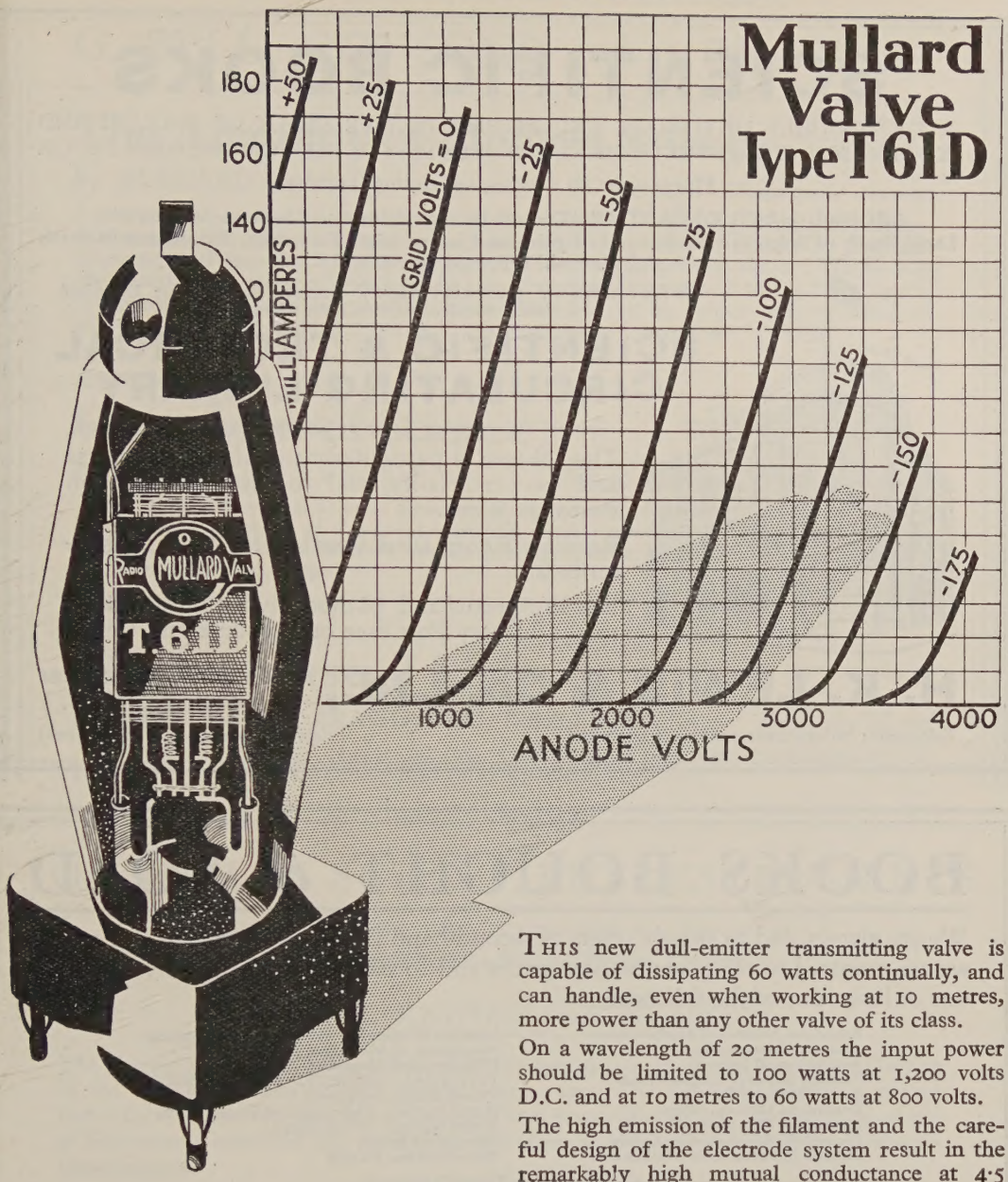
Dubilier Metallized Resistances are sensibly non-inductive at radio frequencies as they consist of a single straight conductor only.

1 WATT 1/-
2 WATT 2/-
3 WATT 3/-

R 7.

DUBILIER CONDENSER CO. (1925) LTD.
DUCON WORKS, VICTORIA ROAD, NORTH ACTON, W.3.

Mullard Valve Type T61D



THIS new dull-emitter transmitting valve is capable of dissipating 60 watts continually, and can handle, even when working at 10 metres, more power than any other valve of its class.

On a wavelength of 20 metres the input power should be limited to 100 watts at 1,200 volts D.C. and at 10 metres to 60 watts at 800 volts.

The high emission of the filament and the careful design of the electrode system result in the remarkably high mutual conductance at 4·5 mA/volt with an amplification factor of 21.

OPERATING DATA

Filament Voltage	6·0 *volts*
Filament Voltage	4·25 *amps.*
Total emission	1,500 *mA. approx.*
Max. Anode Voltage	2,000 *volts D.C.* / 1,500 *volts R.M.S.*
Anode Impedance	4,700 *ohms*
Amplification Factor	21
Anode dissipation (continuous operation)	60 *watts*

PRICE £6. 10s.

Mullard

THE·MASTER·VALVE

MADE IN ENGLAND

ADVT. *The Mullard Wireless Service Co., Ltd., Mullard House, Charing Cross Road, London, W.C. 2*

SCIENTIFIC BOOKS

A THOROUGHLY UP-TO-DATE AND REPRESENTATIVE STOCK OF ENGLISH AND FOREIGN SCIENTIFIC AND TECHNICAL WORKS MAY BE SEEN ON THE FIRST FLOOR

Please write for Catalogues, stating interests

SECOND-HAND DEPARTMENT—140 GOWER STREET. Telephone: Museum 4031
Large Stock of Scientific Books: out of print and scarce scientific works. Books not in stock sought for and reported when found.

STATIONERY DEPARTMENT—Scientific and General: Loose-leaf Note Books, Record Cards, Filing Cabinets, etc.

SCIENTIFIC & TECHNICAL CIRCULATING LIBRARY

Annual Subscription: Town or Country from One Guinea

The Library is particularly useful to SOCIETIES and INSTITUTIONS, and to those engaged on Research Work.

Reading Room (first floor) open daily to subscribers.

Full Prospectus and Bi-monthly list of additions to the Library Post Free on Application.

H. K. LEWIS & Co. Ltd., 136 Gower Street London, W.C. 1

Telegrams: 'PUBLICAVIT, EUSROAD, LONDON' *Telephone:* MUSEUM 7756 (3 lines)

BOOKS BOUGHT & SOLD

We are always glad to receive offers of second-hand Books and Journals, for which we are prepared to offer fair market prices and to pay cash; particularly complete sets or long runs of the following:

American Journal of Mathematics
Annali di Matematica
Astrophysical Journal
Faraday Society. Transactions
Fortschritte der Physik
Journal de Mathématiques
Journal de Physique
Journal für Mathematik. Crelle
Journal (Quarterly) of Mathematics
London Mathematical Society. Proceedings
Messenger of Mathematics
Philosophical Magazine
PHYSICAL SOCIETY. PROCEEDINGS
Physikalische Zeitschrift
Royal Society. Philosophical Transactions
Royal Society. Proceedings
Science Abstracts
Zeitschrift für Physik

And many other Scientific Journals

We shall also be glad to have enquiries for both new and second-hand books, of which we carry an immense stock. Librarians supplied with English and Foreign books. Librarians are invited to send us their lists of books wanted.

Catalogue 386, mainly Physics, Chemistry, Mathematics and Publications of the Learned Societies. Post free on application.

W. HEFFER & SONS, LTD, CAMBRIDGE

BOOKSELLERS AND PUBLISHERS

Licensed Valuers for Probate

TELEPHONE: 862 TELEGRAMS AND CABLES: "Heffer Cambridge"

G. BELL AND SONS ★ ★ ★ New Books

The Anatomy of Modern Science

by BERNARD BAVINK *700 pages. Illustrated.* 21*s. net*

A masterly survey and interpretation of the entire range of modern scientific knowledge. Acclaimed of outstanding interest and importance by men of science in England, Germany and America. "He is a man of profound scientific knowledge....The book is characterised by the moderation of its tone, by its clear statement of the facts, by the formulation of a clear case against materialism and by its proper understanding of what the scientific attitude must be." PROF. E. N. DA C. ANDRADE.

Ready in March

Great Men of Science

by DR PHILIPP LENARD *With 70 Portraits. About* 12*s.* 6*d. net*

This important book is at once a history of the development of science and a series of biographical portraits of the great scientific discoverers from Pythagoras to William Crookes; that is to say from the first awakening of the scientific spirit to the beginning of the twentieth century.

Ready Shortly

The Universe of Light

by SIR WILLIAM BRAGG, O.M., K.B.E., F.R.S. *About* 12*s.* 6*d net*

A book for the general reader yet one which will be read with enjoyment by the professed scientist. With many illustrations (several in colour).

JOURNAL *of* SCIENTIFIC INSTRUMENTS

PRODUCED BY THE INSTITUTE OF PHYSICS

With the co-operation of the National Physical Laboratory

SPECIAL EXHIBITION NUMBER

The *Journal of Scientific Instruments* is the only journal in which the Annual Exhibition of the Physical Society is *fully* reported each year. The February issue is a Special Exhibition Number devoted to the Twenty-Third Annual Exhibition. The different classes of instruments exhibited are dealt with in a series of articles by well-known writers on these subjects.

SPECIAL ARTICLES

The *Journal* now publishes each month an article specially written by a recognized authority on the subject. These articles deal with the trend of development of various classes of instruments, and with matters of general interest to all manufacturers.

It is the English journal devoted to scientific instruments, their principles, construction and use, and receives contributions not only from at home but also from abroad.

The subscription price for a yearly volume is 30/- net (post-free) payable in advance; separate parts 2/6 net each (post-free 2/8)

SUBSCRIPTIONS MAY BE SENT TO ANY BOOKSELLER, OR TO THE

CAMBRIDGE UNIVERSITY PRESS, FETTER LANE, LONDON, E.C. 4

THE PHYSICAL SOCIETY

President:

PROFESSOR A. O. RANKINE, O.B.E., D.Sc., F.INST.P.

IT BECOMES yearly more difficult to keep in touch with the growing volume of Physical research: but by means of its publications, its meetings and lectures, and its annual exhibition, the Physical Society does much to keep its members acquainted with modern developments in all branches of Physics. FELLOWSHIP of the Society is open to all who are interested in Physics. Fellows may attend all meetings of the Society, and they receive PUBLICATIONS 1, 2, 3, and 4 below, as issued. Persons under twenty-six years of age may become STUDENT members. Students receive the *Proceedings* and the *Bulletin* and *Agenda Paper*, and may obtain the other publications at special rates.

PUBLICATIONS

1. *The Proceedings of the Physical Society*, published six times annually, containing original papers, lectures by specialists, reports of discussions and of demonstrations, and reviews.

2. *Science Abstracts A*, published monthly, in association with the Institution of Electrical Engineers, covers practically the whole field of contemporary physical research.

3. *The Reports* are critical monographs on special subjects, prepared by experts and issued from time to time by the Society.

4. *The Bulletin* and the *Agenda Paper*, issued fortnightly during the session, keep members informed of the discussion on papers read at the previous meetings, and of the agenda of future meetings and business of the Society generally.

5. *Science Abstracts B* (ordinary subscription 30/- per annum), which covers in Electrical Engineering a field similar to that covered by Science Abstracts A in pure physics, may be obtained by Fellows for a nominal subscription of 10/- per annum.

MEETINGS

The Society holds about fifteen meetings each session, for the reading and discussion of papers, for experimental demonstrations and for special lectures, including the GUTHRIE LECTURE, given annually by a physicist of international reputation, in memory of the founder of the Society. In addition, a Provincial meeting is generally arranged once a year.

THE ANNUAL EXHIBITION

The Annual Exhibition of Scientific Instruments and Apparatus, which is held at the Imperial College in January, is recognised as one of the most important functions of its kind during the year. A descriptive catalogue is issued to members which provides a valuable record of each year's exhibition.

APPLICATIONS FOR MEMBERSHIP

A candidate for election to FELLOWSHIP must as a rule be recommended by three Fellows, to two of whom he is personally known. A candidate for STUDENT membership must be recommended from personal knowledge by a Fellow.

SUBSCRIPTIONS

Fellows pay an Entrance Fee of £1. 1*s*. and an Annual Subscription of £2. 2*s*. Students pay only an Annual Subscription of 10*s*. 6*d*. No entrance fee is payable by a Student on transfer to Fellowship.

Further information may be obtained from the

ASSISTANT SECRETARY

THE PHYSICAL SOCIETY, 1 LOWTHER GARDENS, EXHIBITION ROAD, LONDON, S.W.7

THE PHYSICAL SOCIETY

PHOTO-ELECTRIC CELLS AND THEIR APPLICATIONS

A DISCUSSION ON THE HISTORY, DEVELOPMENT AND MANUFACTURE OF PHOTO-ELECTRIC CELLS, AT A JOINT MEETING OF THE PHYSICAL AND OPTICAL SOCIETIES

Editor:

JOHN S. ANDERSON, M.A., D.Sc., F.Inst.P.

Containing 31 papers on the Theory, Manufacture, Properties and Applications of Photo-Electric and Selenium Cells, together with a Report of the Discussion thereon

236 pages, with numerous illustrations

12*s.* 6*d.*; 13*s.* 0*d. post free*

DISCUSSION ON AUDITION

Containing 18 papers. 151 pages, with numerous illustrations

7*s.* 0*d.*; 7*s.* 3*d. post free*

THE TEACHING OF GEOMETRICAL OPTICS

84 pages. Containing 15 contributed at a meeting of the Society

4*s.* 6*d.*; 4*s.* 8*d. post free*

PUBLISHED BY

THE PHYSICAL SOCIETY

1 LOWTHER GARDENS, EXHIBITION ROAD

LONDON, S.W.7

CAMBRIDGE UNIVERSITY PRESS

Conduction of Electricity through Gases

VOLUME II

By Sir J. J. THOMSON and G. P. THOMSON

Demy 8vo. With a frontispiece, 5 plates, and 232 text-figures. 30*s. net*

This volume completes the third edition of *Conduction of Electricity through Gases*. It deals with ionisation by collision and by X-rays and with the properties of the electric discharge in all its forms; glow, spark and arc. The progress of the subject since the last edition has made it necessary to re-write almost the whole of this part of the book and to expand it to about twice its former size.

The Form and Properties of Crystals

AN INTRODUCTION TO THE STUDY OF MINERALS AND THE USE OF THE PETROLOGICAL MICROSCOPE

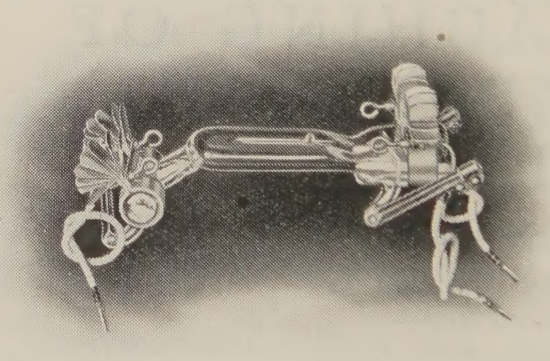
By A. B. DALE

Crown 8vo. With 124 text-figures, etc. 6*s. net*

Difficulty has been experienced in finding a text-book simple enough for those with only an elementary knowledge of physics and mathematics, comprehensive enough to give them most of what they need in a volume of moderate size, and at the same time inexpensive enough to suit their purses. This text-book represents an effort to supply their needs.

VITREOSIL

(PURE FUSED QUARTZ OR SILICA)



In addition to the mercury arc burners such as are illustrated and which are used for the irradiation and examination of materials, we construct burners for special work to customers' own requirements.

We shall be glad to receive enquiries from those who are engaged in work which requires either special burners or special apparatus which must transmit ultra-violet radiation.

SOLE MANUFACTURERS:

THE THERMAL SYNDICATE LTD

WALLSEND-ON-TYNE

Established over a quarter of a century

London Depot: THERMAL HOUSE, OLD PYE STREET, S.W.1

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 45, PART 2 *March* 1, 1933 NO. 247

621.317.755

TIME-MARKING A CATHODE-RAY OSCILLOGRAM

BY LEWIS F. RICHARDSON, F.R.S.

Received August 2, 1932. *Read December* 2, 1932.

ABSTRACT. Time-marks have been arranged as little blurs or gaps in the trace, by periodically unfocusing the electron stream. The current in and voltage across a conductor can thus be recorded simultaneously with the time on a single oscillogram.

ALTHOUGH excellent arrangements for recording the time on an oscillogram are already available, such as the linear time-base used by Appleton, Watson Watt and Herd* combined with the trigger of Finch, Sutton and Tooke†, yet these give the time as one co-ordinate and so leave only one co-ordinate for the phenomenon to be investigated. There is an obvious need for a different method, now to be described, having the advantage that both co-ordinates are allocated to the phenomena, say one co-ordinate for voltage, the other for current. It would be possible to distort the trace by an impressed ripple. In particular D. W. Dye‡ has shown how to apply a small deflection either parallel to a fixed direction or radial or circular. But ripples are not satisfactory for two reasons: (i) the distortion hinders measurement; (ii) the ripple is invisible where its displacement is parallel to the undisturbed trace. Instead, time-marks at intervals of 10^{-3} or 10^{-4} sec. have been arranged as little blurs or gaps in a trace otherwise undistorted.

In recent types of oscillograph (I have used that of Standard Telephones and Cables Ltd.) the spot is focussed by applying a suitable potential to a cylindrical shield surrounding the filament. The time-marks have been made by unfocusing the spot at regular short intervals of time, by applying to the shield a periodically varying potential generated by a triode oscillator of an ordinary kind§.

* *Proc. R. S.* A, **111**, 672 (1926).
† *Proc. Phys. Soc.* **43**, 502 (1931).
‡ *National Physical Laboratory; Report for the year* 1924, pp. 77, 78.
§ Like figure 177 of L. B. Turner's *Wireless*, p. 267 (Camb. Univ. Press, 1931).

The appearances are shown in oscillograms 1, 2, 3, 4, in all of which the two co-ordinates of the phenomenon depend severally on the voltage and current in a condenser supplied with a.c. at 50 ~ from the town mains.

The tuning capacitance of the time-marker was placed close to the oscillograph so that the observer could vary the frequency of the time-marks while watching the oscillograph trace. But the tuning inductance was put 5 metres away to avoid its stray field.

The effects of mistuning are like those seen on a stroboscope. For example suppose that by means of a frequency, known to be about 20 times that of the mains, 20 stationary time-marks appear on the orbit of the light-spot; and that the tuning capacitance of the time-marker is then gradually decreased until 21 stationary marks appear. During this passage various appearances are seen: time-marks hurrying to and fro, and blurred stationary patterns of about $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$, ... of the spacing characteristic of the 20th harmonic. It is easy to recognize these intermediate patterns by their closeness and by their blurring, and so to avoid them. They have been avoided in the present work.

To decide which way the light-spot was moving, the following experiment was made. The tuning capacitance of the time-marker was set to give stationary harmonic marks. The capacitance was then suddenly increased by a small fraction of the change that would produce the next lower harmonic. This caused the time-marks to move along the trace in the same sense as the light-spot. Thus we can put the head on time's arrow.

Some puzzling oscillograms concerning neon-lamps were resolved by the aid of slowly moving time-marks. A piece of trace that had appeared to be single was seen to be traversed by two sets of time-marks moving opposite ways.

The following three propositions have been found very serviceable in the interpretation of oscillograms:

I. The speed of the moving time-marks is the same fraction of the speed of the light-spot in all parts of its trace.

II. When the marks are stationary, a sudden increase of the tuning capacitance by one five-hundredth of its effective amount, including self-capacitance, causes the marks to advance with one-thousandth* of the speed of the light-spot.

III. The separation of sufficiently close time-marks is to their speed in a ratio which is the same for all parts of the trace.

Proofs of these propositions are given in the Appendix.

As this scheme has been tried out in various special forms, some of which gave neat definite gaps in the trace while others gave blurred patches too vague for accurate measurement, it may be desirable to describe the arrangements in more detail.

Figure 1 shows a simple connexion. On the left F is the filament, S the shield, and A the anode of the oscillograph. The focus is adjusted by the voltage B of the

* More accurately 1/1001. [Note added December 3, 1932.]

No. 1. 50 × 20 marks per second. Circuit of figure 1

No. 2. 50 × 20 marks per second. Circuit of figure 2; $W \rightarrow \infty$, $R = 0{\cdot}28$ MΩ., $M = 129$ V., $K = 50\,\mu\mu$F.

No. 3. 50 × 86 marks per second. Circuit rather like that of figure 2. Undesirable ripple.

No. 4. 50 × 107 marks per second. Circuit of figure 1. Separate beads, easily counted.

To face page 137

battery. There was a range of about 12 volts over which the focus was moderately good; so the oscillating voltage supplied by the coil L had to be of this order. If B were set for the middle of the range of fair focus and the amplitude ϕ of the time-marking oscillation were large enough, we should expect a blur for every half wave. Actually ϕ was usually too small for that, so B was set near one end of the range, making only one blur in each whole wave. For example in oscillogram No. 1, the phenomenon had a period of 0·02 sec., the time-marker a period of 0·001 sec., and there are 20, not 40, gaps.

It is an advantage to have the resistance R adjustable between 10^4 and 10^7 Ω, for by its aid the conspicuousness of the time-marks is easily controlled. The trade-name of the resistance was "power-clarostat, universal range."

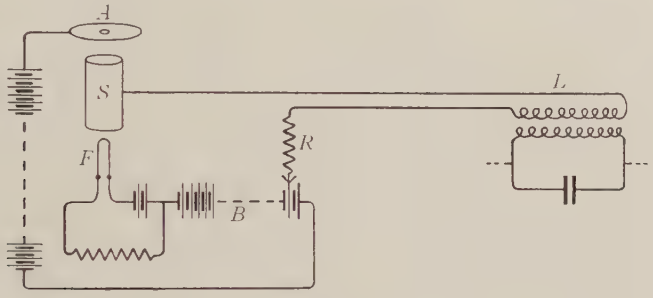


Figure 1.

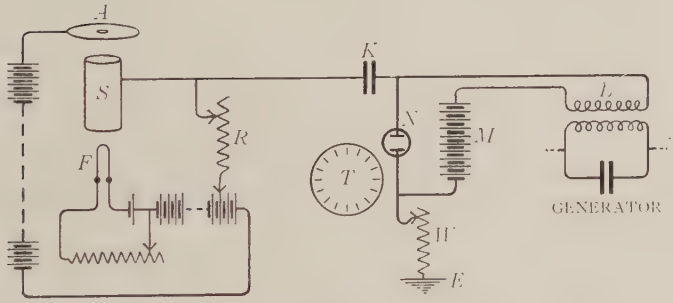


Figure 2.

In oscillogram No. 1 the gaps are too large and the trace near the gaps too distorted. The circuit of figure 2 gave better results.

Here N was a neon (osglim) lamp to make sudden jerks of potential by lighting 10^3 times per second. As the oscillator alone did not give enough volts to flash the lamp, the battery M was added. But even so the range from maximum to minimum given by the coil L had to exceed some 40 volts in order both to light and extinguish the lamp. The function of the capacitance K of about 0·001μF is to transmit the jerk but to insulate for steady currents. If the potential at one side of N be steadied by shorting W to earth, the jerk of potential will all come on the other side of N.

When this was tried, the oscillograph emitted a note of the frequency of the time-marker, and lest the filament should be broken, the circuit was hastily disconnected. But a moderate resistance W might be helpful. As the jerk at the lighting of the osglim only lasts about 10^{-5} sec., it seemed desirable to treat the connexion from N to S in the manner customary for radio-frequency, taking care to avoid stray capacitances to earth and removing the choke-coil which the makers put in the cap of the osglim. No protective coil was needed, as the power supplied by L was insufficient to cause arcing. Oscillogram No. 2 was marked by the circuit of figure 2. It is seen that the marks are definite to 3×10^{-5} sec.

Photography was difficult because the marks were seldom still for 3 seconds continuously. Their random to-and-fro motion along the trace was presumably due to variations in the frequency of the Paisley town mains; and this disturbance was less in the small hours of the morning. By watching the trace through a magnifying glass and beginning the exposure when the marks were still, relative to the crystals on the screen, I obtained about one successful photograph for five exposures. By a fortunate chance a cinematograph lens was found which had a field curved so as approximately to fit the end of the oscillograph like a cap. The curvature was therefore opposite to that given by a simple bi-convex lens. This lens is marked "$3\frac{1}{2}$ in. Paragon Special," and is a variety of the Petzeval objective*. Ilford "Golden Iso-Zenith" plates were used with satisfaction. Almost confluent marks on the negative were sometimes separated by making a contact print on a process plate.

Standards of frequency. The triode oscillator kept a frequency sufficiently steady for photography; for as tested by beats against a vibrating steel bar, its frequency varied only by 1 part in 2000 during the course of an evening. But variations of several per cent from month to month, depending on the filament-temperature and on the potentials of the anode and grid, made it desirable to refer to better standards. These are well known, and include for instance tuning-forks, Campbell hummer-bars and wave-meters†. But perhaps I may mention that a gramophone, a protractor and a neon lamp make a domestic stroboscope; also that a water-tube resonator with a telephone fixed near its mouth is a cheap and convenient instrument.

Peculiarities at 10,000 ~. The audio-frequency oscillator would not go so high until provided with coils of less self-capacitance. Then small time-marks got across mysteriously from oscillator to oscillograph even when there was no wire connexion. These marks however were not of much use, because they were small displacements of the trace all parallel to one direction and therefore invisible for those parts of the trace that went in that direction when unmarked. The circuit of figure 1 provided proper time-marks visible in all parts of the trace. The circuit of figure 2 did not succeed any better, owing to insufficient voltage to flash and extinguish the osglim lamp.

* J. W. French, *Dict. App. Phys.* **4**, 158.

† D. W. Dye, "Radio-frequency Measurements," *Dict. App. Phys.*

Example of an observation to standardize a wave-meter:

Time of day a.m.	All capacitances mentioned below were those of the generator; and $C = 0{\cdot}066_0\,\mu$F.
2.55	Stroboscope of 100 holes illuminated by neon lamp at the town's frequency, 100 revolutions in $100{\cdot}1_0$ sec.
2.59	Wave-meter maximum resonance at $C + 0{\cdot}00036$
3.10	Photograph of oscillogram at $C + 0{\cdot}00016$
3.15	Wave-meter maximum at $C + 0{\cdot}000275$
3.20	Stroboscope as above, in $100{\cdot}0_6$ sec.
3.30	A correction for the effect of the galvanometer lamp on the filament-current of the generator was easily determined by holding the same harmonic pattern of time-marks steady on the oscillogram while switching the lamp on and off. It was found that the capacitance at the time of the photograph should be increased to $C + 0{\cdot}00036$, leaving the other capacitances as stated

The change of frequency ν with capacitance C, although given roughly by $\nu = 1/2\pi\sqrt{(LC)}$, is easily and more accurately determined by suddenly turning the knob of an air condenser so as to pass from the photographed harmonic pattern to the next, a change of 50 $\sim$. In this way it was found that 50 $\sim$ corresponded to $0{\cdot}00059\mu$F. On projecting the photograph to form an image 44 cm. in diameter it was just possible to count the time-marks. There were 196 of them.

The frequency of the wave-meter was accordingly calculated thus:

$$\frac{100 \times 100}{100{\cdot}1_0 \div 100{\cdot}0_6} \times 196 - \left(\frac{36 + 27{\cdot}5}{2} - 36\right)\frac{50}{59} = 9792 + 4 = (9796 \pm 10) \sim,$$

an absolute determination. Another set of observations gave 9801 $\sim$. The separation of these marks is about equal to the diameter of the light-spot, so that closer marks could hardly be resolved and there is little prospect of counting 400 time-marks on a single oscillogram.

ACKNOWLEDGMENT

The Government Grant Committee of the Royal Society provided some apparatus, which has been a great help.

APPENDIX

Proofs of theorems

Let T denote the time taken by the light-spot to make each circuit of its trace. T

Let $(T + \epsilon)/n$ denote the interval of time between successive time-marks; and let n be an integer. ϵ, n

If $\epsilon = 0$ the marks are stationary. Let us ignore in this theory those rather blurred stationary marks for which n is an integer plus a simple fraction such as $\frac{1}{3}$, $\frac{1}{2}$, $\frac{2}{3}$; for it has been shown how to observe and reject such marks experimentally.

The arcual distance S_1 between two successive stationary marks is travelled by the light-spot in a time T/n. So if V_1 denote the mean speed of the light-spot for that part of its course, S_1 V_1

$$TV_1 = nS_1 \qquad \text{.....(1).}$$

But if $\epsilon > 0$, consider any time-mark. Call it the zeroth mark occurring at time zero. The light-spot returns to the same point on the screen at the time T. But the nth mark has not yet arrived and so the light-spot moves on through a further
S distance S, say, to where the nth mark occurs at the time $T - \epsilon$. If S is a small fraction of the distance between successive marks, we see an apparent forward
U motion, familiar on the stroboscope. Let U be the speed of this apparent motion. It is due to an advance S during the time between the zeroth and the nth mark. This time-interval is $T + \epsilon$. So

$$S = U(T + \epsilon) \qquad \text{.....(2).}$$

V Let V denote the speed of the light-spot for the arc S. Then

$$S = V.\epsilon \qquad \text{.....(3).}$$

Here S is so small that V is indistinguishable from the *instantaneous* speed of the light-spot.

Eliminating S between (2) and (3),

$$U = V\frac{\epsilon}{T + \epsilon} \qquad \text{.....(4).}$$

Now $\epsilon/(T + \epsilon)$ is independent of position on the trace and so theorem I on page 136 follows.

Next suppose that when the marks are stationary ($\epsilon = 0$) we suddenly increase
dC the tuning capacitance C of the time-marker by an amount dC so that dC C is small. Assuming for simplicity that the period P of the time-marker is given by the usual formula

$$P = 2\pi\{L(C + \gamma)\}^{\frac{1}{2}} \qquad \text{.....(5),}$$

L, γ in which L is the fixed tuning inductance and γ its self-capacitance. Differentiating (5) logarithmically, we have

$$\frac{dP}{P} = \frac{1}{2}\frac{dC}{C + \gamma} \qquad \text{.....(6).}$$

But

$$\frac{dP}{P} = \left(\frac{T + \epsilon}{n} - \frac{T}{n}\right)\Big/\left(\frac{T}{n}\right) = \frac{\epsilon}{T} \qquad \text{.....(7).}$$

Eliminating dP/P between (6) and (7) and inserting the result in (4),

$$U = V\frac{dC}{2(C + \gamma) + dC} \qquad \text{.....(8)*,}$$

which includes theorem II.

To prove theorem III we compare equation (1) with equation (4), obtaining

$$S_1 = \frac{T}{n}V_1 = \frac{T}{n}V.\frac{V_1}{V} = \frac{T}{n}U\left(\frac{T + \epsilon}{\epsilon}\right)\frac{V_1}{V} \qquad \text{.....(9).}$$

Now as $n \to \infty$, $V_1/V \to 1$, leaving

$$S_1 = U.T\frac{T + \epsilon}{\epsilon n} \qquad \text{.....(10),}$$

* Revised December 3, 1932.

and although we cannot make n approach ∞ in the experiments, yet we can usually make n large enough for (10) to be an excellent approximation. The factor $T(T+\epsilon)/\epsilon n$ is independent of position on the oscillogram; whence theorem III.

DISCUSSION

Prof. G. I. Finch. The trace-luminosity afforded by a medium-speed electron beam focussed by a Wehnelt cylinder is often so low that the phenomenon under observation must be repeated many times in order to obtain a photographic record. The value of Dr Richardson's neat and useful method of time-marking would therefore be greatly enhanced if means could be found of locking time-marking and trace together in such a manner that the time-marks coincided accurately throughout a large number of repeated traces.

Author's reply. Thinking of the work of Finch, Sutton and Tooke on synchronization I made many attempts to use a continuous method of synchronization, namely to take as the time-marker a high harmonic of the trace-frequency. The harmonic was produced by a copper-oxide rectifier and picked out by a resonant triode circuit. This scheme succeeded with the lower harmonics such as the 10th, but was no use at all with the higher harmonics such as the 200th.

532.13:541.182.62

THE FLOW OF LIQUID SUSPENSIONS

By E. G. RICHARDSON, B.A., Ph.D., D.Sc.
AND
E. TYLER, D.Sc., F.Inst.P.

Received October 13, 1932. *Read December* 16, 1932.

ABSTRACT. Measurements of the velocity from point to point in a liquid rotating in the space between two concentric cylinders, of which the outer rotates while the inner is still, are made by means of a hot-wire anemometer. It is shown that in a suspension the velocity-gradients are abnormal but can be explained in terms of a variable viscosity, which is a function of the velocity-gradient.

§ 1. INTRODUCTION

The anomalous viscosity of disperse systems has become a common subject of research in recent years. As soon as liquid suspensions and gels were investigated in viscometers of either the capillary-flow or the concentric-cylinder type, it was found that the coefficient of viscosity was not a constant for a given liquid; for instance, that in the capillary-flow apparatus the pressure-gradient increased less rapidly than the outflow it produced, or, that in the concentric-cylinder apparatus the velocity of the outer rotating cylinder increased more rapidly than the torque it produced on the inner. It was, in fact, impossible to measure the viscosity absolutely, the measured value being dependent on the apparatus and the conditions under which it was used. In order to obtain data for comparing the behaviour of such systems with that of ordinary liquids, it is necessary, as Hatschek* points out, to measure the velocity from point to point across the liquid, and to compare velocity-gradients.

Two attempts in this direction have recently been made, both with a flow apparatus and with a gelatine solution. Kroepelin† traversed a Pitot tube across the end of the flow tube at the place where it debouched upon the outflow reservoir, and obtained curves of velocity transverse to the flow, showing small departures from the Poiseuille (parabolic) distribution. The sol was a gelatine solution. Pichot and Dupin‡ used a fluid partaking of both suspension and sol characteristics, consisting as it did of a 7 per cent solution of gelatine in water, with aluminium particles to make visible the relative velocities of different parts when an exposure was made. From the two photographs given (water and gelatine solution respectively) it is difficult to detect any considerable difference in the behaviours of homogeneous and heterogeneous liquids. We have found that gelatine is very difficult to work with, owing to the gradual change in its properties which occurs during an experiment.

* *Viscosity of Liquids*, bibliography given on p. 227 (1928).
† *Z. f. phys. chem.* **149**, 291 (1930).
‡ *Comptes Rendus*, **192**, 1079 (1931).

§ 2. EXPERIMENTAL ARRANGEMENTS

The hot-wire anemometer can readily be adapted to measurements in liquids*; there is however a smaller change of resistance with velocity of the liquid than when the same wire is used under the same conditions in air.

In the present research a vertical wire of nickel, 0·001 in. in diameter, was used. It was heated by a current of 0·30 A. and was calibrated by being mounted on a small whirling arm which carried it in a circular path, 42 cm. in diameter, through the liquid contained in an annular tank. The resistance of the wire was measured in a Kelvin bridge circuit for various speeds of revolution of the whirling arm, an allowance being made for the drift of the fluid in the tank. Typical calibration curves are shown on figure 1.

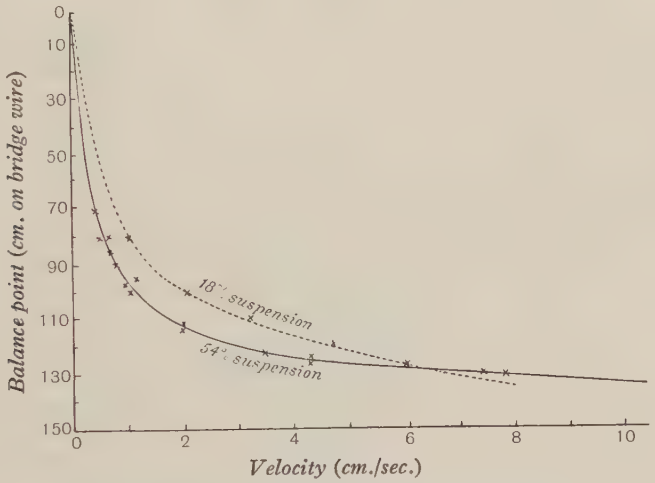


Figure 1. Calibration curves for hot-wire anemometer.

Before applying these curves, another test is necessary. When the wire is used in close proximity to a solid surface, e.g. the walls of the vessel containing the fluid in motion, there is a loss of heat due to conduction across the thin film of fluid between the wire and the surface. This is additional to the heat-loss by convection due to the moving fluid. In this case, it was found, by moving the wire towards the outer cylinder while the liquid inside was still, that the resulting decrease of resistance did not arise until the wire was within 0·5 mm. of the brass surface. As all measurements of flow were made outside this distance, it was unnecessary to apply a correction.

The actual apparatus for the viscosity and velocity gradients was of the Couette type, and is drawn in figure 2; it had been previously used in a slightly different form by one of us† for measurements in homogeneous liquids. The hot wire *W*, 1·5 cm. long, is carried vertically on the ends of two fine needles inserted through

* Davis, *Phil. Mag.* **40**, 692 (1920) *et seq.*

† *J. Sci. Inst.* **6**, 337 (1929).

ebonite bushes in the outer brass cylinder, which is rotated by the pulley *P* below. Electrical connexion to the fork blades is assured by dippers working in the fixed annular rings of mercury *GG* mounted below the fork. The speed of this cylinder is observed by means of a neon lamp *N* illuminating the stroboscopic disc *D*, and the flashing of the lamp is controlled by a vibrating reed of standardized frequency.

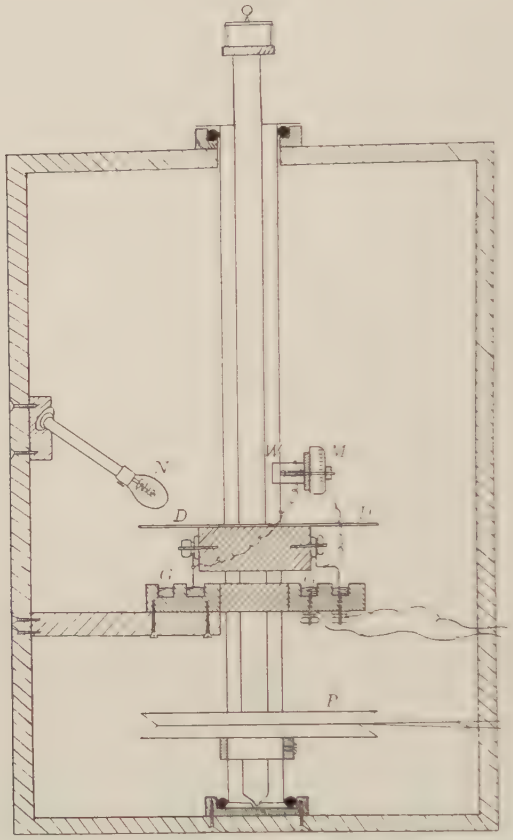


Figure 2. Apparatus.

The hot wire can be traversed across the annular space between the two cylinders by means of a micrometer head *M*. The cylinders are 1 in. and $\frac{1}{2}$ in. in diameter respectively and 1 ft. long. The torque communicated to the inner cylinder is taken up by a suspension wire, not shown in the sketch, the twist on which was indicated by mirror, lamp and scale. With the outer cylinder rotating at constant speed, measurements of the resistance of the wire at various points across the inter-cylindrical space were made, and these were converted into relative velocities by the use of the appropriate calibration curve. Figure 3 shows a set of velocity-distribution curves for paraffin. These are nearly linear, in accordance with theory. Figure 4 shows graphs of torque (in cm. of scale deflection) against velocity of outer cylinder, similar to those by which anomalous viscosity was first indicated.

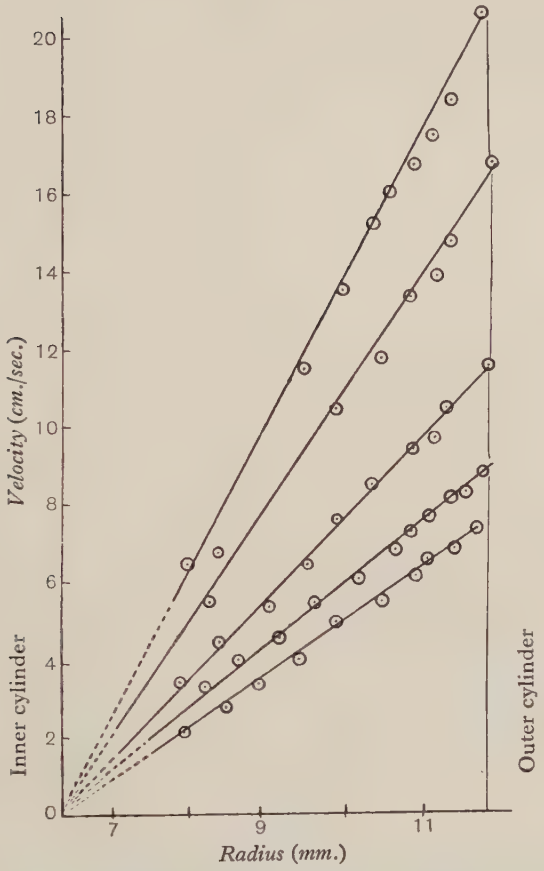


Figure 3. Velocity-gradients in paraffin.

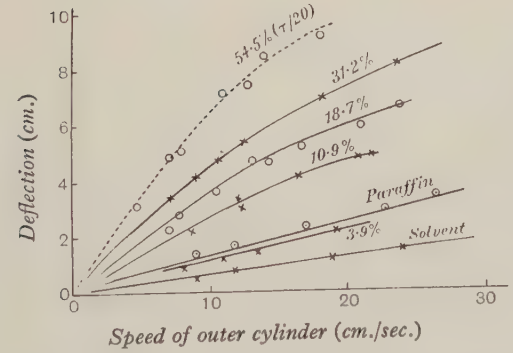


Figure 4. Variation of torque with speed of outer cylinder. The ordinates for the top curve have been reduced to $\frac{1}{20}$.

The paraffin shows, of course, constant viscosity (torque proportional to rate of shear), but the suspensions, with the exception of the weakest, all show diminishing viscosity as the rate of shear is increased. These suspensions were of rice starch in carbon tetrachloride and paraffin, made up according to the recipe of Humphrey and Hatschek* as follows: finely ground rice starch was heated carefully in a steam oven for 2 or 3 hours, then passed through a fine sieve and allowed to cool. The

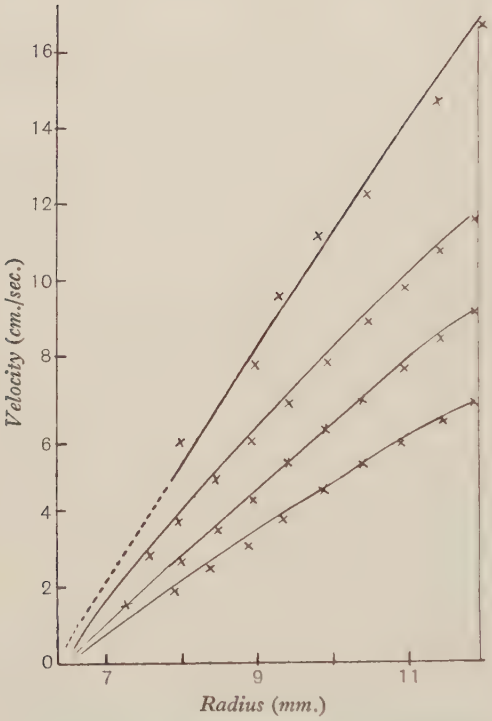


Figure 5. Velocity-gradients in 3·9-per-cent suspension.

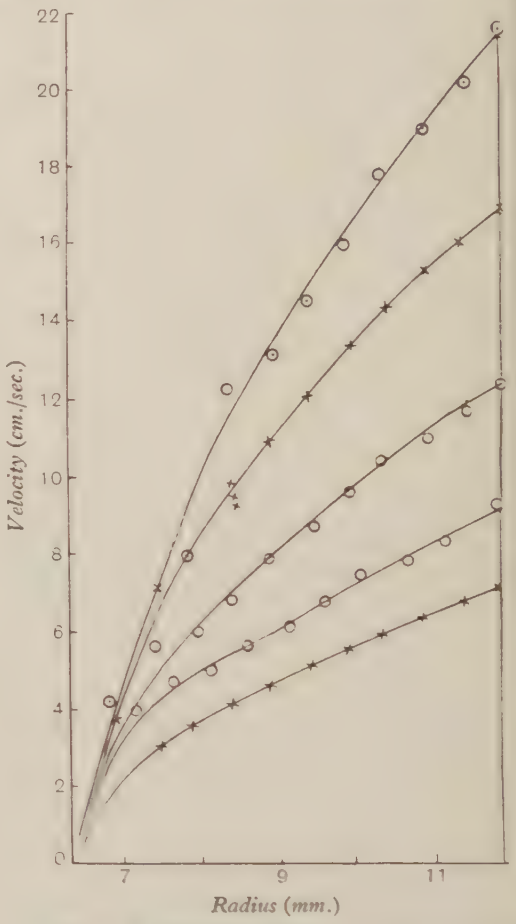


Figure 6. Velocity-gradients in 10·9-per-cent suspension.

suspension was then made up to the required strength with a mixture of paraffin and carbon tetrachloride of density 1·5 gm./cm³. The most concentrated suspension had the consistency of paste, and was chosen to compare with the measurements of Keen and Scott Blair† on clay pastes. The outer annuli of this paste rotate with the outer cylinder as a rigid whole, particularly when the speed of the latter is low.

* *Proc. Phys. Soc.* **28**, 276 (1916). We wish to thank Mr Hatschek for the recommendation to use these suspensions on account of their stability.

† *J. Agric. Soc.* **19**, 684 (1929).

In no case was the critical velocity for turbulent flow exceeded, although there was difficulty in obtaining the torque-measurements for the more concentrated specimens owing to instability of the inner cylinder. A small difference in the height of the column of liquid, made with the object of detecting an end effect, did not alter the velocity-gradients but affected the torque-measurements in the proportion of the wetted surface. Results, the suspensions of four different concentrations, are shown in figures 5, 6, 7, 8. They all, with the exception of the weakest, exhibit gradients varying notably in the neighbourhood of the inner cylinder.

§ 3. THEORY

Theories of colloid viscosity may be grouped under two types:

(i) The hypothesis of structural friction. Buckingham* supposed that, in addition to a viscous force proportional to the velocity-gradient dv/dy between neighbouring annuli, there was a constant force a due to friction between the disperse phase and the neighbouring liquid. Newton's law then becomes

v, y a

$$F = (a - \eta\, dv/dy)\, dS \qquad \text{.....(1)},$$

η being the coefficient of viscosity and dS an element of area. This theory predicts an outer section of fluid endowed with a constant angular velocity in the case of the concentric-cylinder apparatus, moving *en bloc* with the outer cylinder. This theory Pichot† thinks to be confirmed by his measurements with gelatine in a flow tube. The theory also predicts a limiting shear, below which no flow should take place. This limiting stress has unfortunately never been observed.

η, dS

(ii) The hypothesis that the viscosity-coefficient is variable and a function of the velocity-gradient. Porter and Rao‡ write

$$\eta = \eta_0\, (dv/dy)^{-n}.$$

n

For the heterogeneous liquids, for which n is given a small positive value while it is zero for the homogeneous liquids, velocity-gradients similar to those given by the first theory can be constructed. This theory has the advantage that no static friction is expected of the fluid. Newton's law, on this theory, becomes

$$-F = \eta_0\, (dv/dy)^{1-n}\, dS \qquad \text{.....(2)}.$$

In the case of the concentric cylinder, the velocity-gradient at any radial distance r being $r\,d\omega/dr$, the moment of the annulus about the axis is $2\pi r\eta\, (r\,d\omega/dr)\, r$ per unit length, and this moment is the same for every value of r in steady motion, so that the torque τ is proportional to $r^3\eta\, d\omega/dr$. Thus $\tau\, (r^3 d\omega/dr)^{-1}$ should be a measure of relative viscosity. Accordingly, for each of the suspensions of figures 6, 7, 8, the slope of the *angular-velocity* curves was first plotted against r, and hence $\tau\, (r^3\, d\omega/dr)^{-1}$ could be plotted against $r\, d\omega/dr$, the appropriate torque being read from figure 4. The resulting (viscosity, velocity-gradient) curves are shown on

r, ω τ

* *Proc. Amer. Soc. Tent Materials*, **21**, 1154 (1921).
† *J. de Physique*, **3**, 205 (1932).
‡ *Trans. Farad. Soc.* **23**, 311 (1927).

figure 9, the pure solvent being included, for purposes of comparison, at its absolute value ($\eta = 0{\cdot}012$).

Each suspension shows a diminishing viscosity as the velocity-gradient increases, tending however to an asymptotic value at high velocity-gradients. The flow is in fact nearly normal at high velocity-gradients, unless turbulence intervenes. A note-

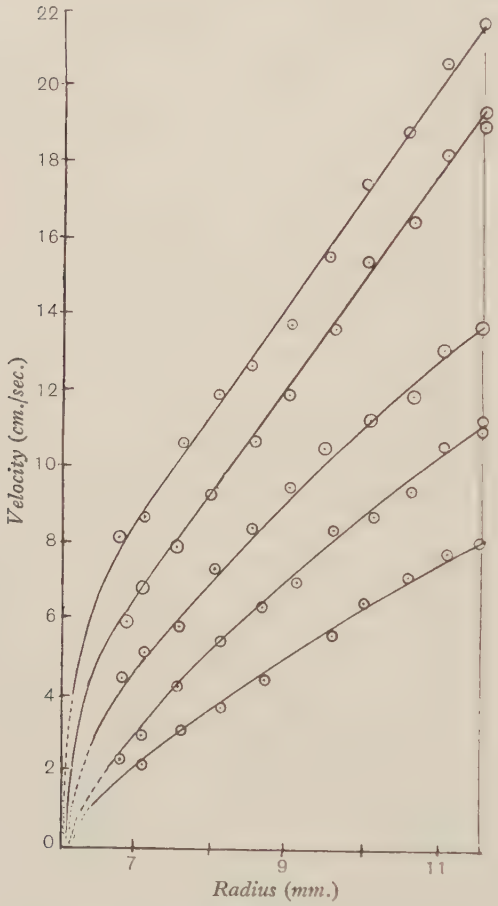


Figure 7. Velocity-gradients in 31·2-per-cent suspension.

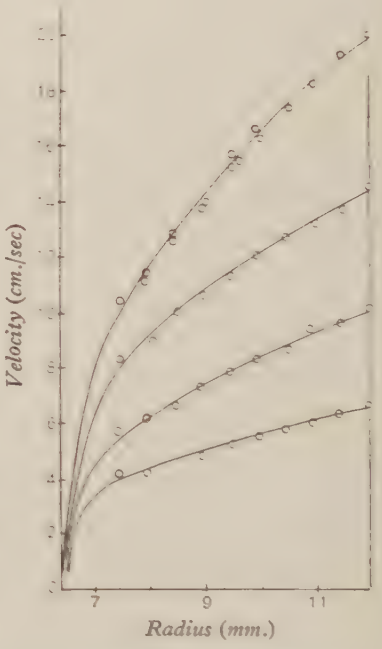


Figure 8. Velocity-gradients in 54·5-per-cent suspension.

worthy example of this normal flow occurs in the top curve of figure 6. By plotting $\log \eta$ against $\log (r\,d\omega/dr)$ the mean values of n were found to be 0·55 and 0·75 respectively for the two medium suspensions, and 0·9 for the paste. These values are of the same order as those deduced by Farrow, Lowe and Neale* by semi-empirical methods from outflow measurements in capillary tubes. An exponential formula such as that suggested by Reiner and Riwlin† might also be fitted to the curves. Finally,

* *J. Test Institute*, **19**, T 18 (1928).
† *Kolloid. Z.* **43**, 72 (1927); **44**, 9 (1928).

in figure 10 we have plotted η against concentration by volume for several values of velocity-gradient, including data for an 18-per-cent suspension for which the curves are not reproduced. The deduction of Einstein* from Stokes's law, that the

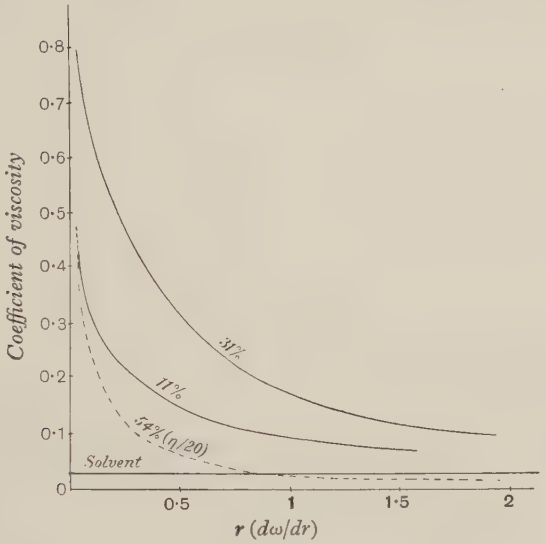


Figure 9. Variation of viscosity with velocity-gradient.

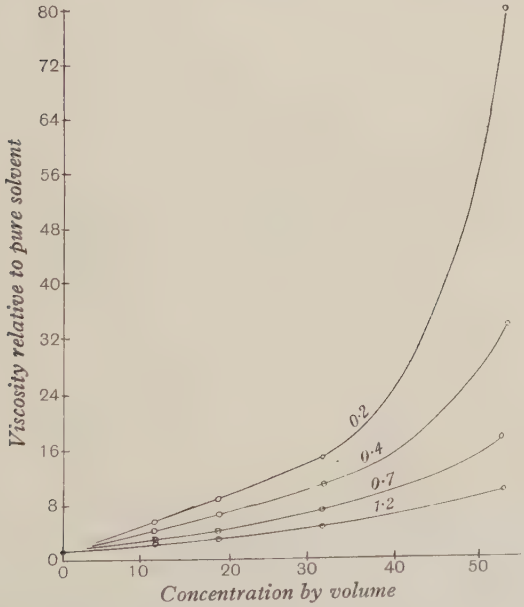


Figure 10. Variation of viscosity with concentration. The numbers against the curves denote $r(d\omega/dr)$.

viscosity should be a linear function of the concentration, fails at large concentrations, but at high velocity-gradients the variation of η with concentration approximates to that of an ideal suspension, conforming to Newton's and Einstein's laws.

* *Ann. der Phys.* **19**, 289 (1906); cf. Humphrey and Hatschek, *loc. cit.*

§ 4. CONCLUSIONS

These results, which we believe to be the first systematic measurement of velocity from point to point in a series of colloids in motion, confirm the earlier deductions from integrated flow-measurements that the coefficient of viscosity in a disperse system diminishes with increasing velocity-gradient, but tends to a constant value at high rates of shear. At moderate rates of shear the distribution of flow in the concentric-cylinder system is such that the velocity falls gradually from the outer cylinder inwards, with a more rapid drop at the surface of the inner cylinder. We have indeed assumed the condition of no slip at the solid boundary. That there is none at the outer cylinder, at least, is shown by the velocity graphs, and our extrapolation to zero velocity at the surface of the inner cylinder is confirmed by the torque on the latter, for in each of figures 6, 7 and 8, the ratio torque velocity-gradient at the inner cylinder is a constant for a particular suspension, although the ratio torque/velocity of the outer cylinder is not. Neither is there sure evidence of a yield value, such as the Buckingham theory requires*. The smallest shear which we applied started the fluid into motion. In this theory the force a, equation (1), which many writers identify with a solid-liquid friction, need not indeed be constant; it can be a function of the velocity-gradient, vanishing with the flow. With this modification, indeed, the two theories coalesce.

As to the ultimate cause of anomalous viscosity, these experiments tell little, except that there is some sort of interference to the flow on the part of particles of the disperse phase, which is reduced as their relative velocity is increased. This may be due to the formation of aggregates gradually broken down as the rate of shear increases, or, as Hatschek† has suggested, to combination of the two phases, the particles so produced being reduced in size as their liquid envelopes are shorn off. That some form of interlocking may well be a property of even a homogeneous liquid is shown by the success which Andrade‡ has had with a formula, based on his concept of "transitory and fluctuating crystallization," which covers the variation of viscosity with both pressure and temperature in a homogeneous liquid.

DISCUSSION

Mr E. W. Madge. In connexion with the Einstein relation

$$\eta/\eta_{\substack{\text{dispersion}\\\text{medium}}} = 1 + K\theta,$$

where θ is the volume concentration, it would be of value if the authors would state what in their opinion is the most probable value of K over the linear portion of the curve. Such a statement based on what appears to be an accurable method of investigating the flow of suspensions would no doubt serve to settle the differences

* Except in the case of the thick paste.

† *Kolloid. Z.* **40**, 53 (1926).

‡ *Nature*, **125**, 309, 580 (1930).

of opinion concerning the hypotheses on the basis of which such a formula is derived.

The linear formula also implies that the viscosity relative to the dispersion medium depends only on the total volume of the suspended particles and not on their actual dimensions. It would be interesting to know whether the authors have in mind extending their work to determine over what range this is valid.

Authors' reply. Owing to the dependence of η on velocity-gradient even at low concentrations, it is rather difficult to deduce a precise value for K from the experiments. If we understand Einstein's formula correctly, it depends on the laws of classical hydrodynamics and assumes no material interference between the suspended particles or adsorption of the dispersion medium. The existence of anomalous viscosity seems to indicate that such interference does in fact take place. If this be so, the application of Einstein's formula to suspensions seems of doubtful validity, but we agree that an investigation of the effect of the size of the particles would be useful. However, on the assumption that the interference is least when the rate of shear is large, the flow then approaching most closely that of a homogeneous fluid, the value of K in the experiments was of the order of that (2·5) deduced theoretically by Einstein.

551.594.12:551.51.051

THE INFLUENCE OF ATMOSPHERIC SUSPENSOIDS UPON THE EARTH'S ELECTRIC FIELD AS INDICATED BY OBSERVATIONS AT KEW OBSERVATORY

By H. L. WRIGHT, M.A.

Communicated by Dr F. J. W. Whipple, October 7, 1932. Read December 2, 1932.

ABSTRACT. Changes in the electrical resistivity of the air and in the potential-gradient may be associated with changes in the concentration of gross particles and of condensation nuclei. Observations of these four quantities made at Kew Observatory over a period of three years are analysed with a view to formulating functional relations between them. The observations indicate that resistivity depends upon the number of gross particles and of condensation nuclei, while potential-gradient varies with resistivity and the number of nuclei. The theoretical implications of this statement are considered.

§ 1. INTRODUCTION

Our knowledge of the electrical conditions of the atmosphere is derived very largely from observations of the potential-gradient near the earth's surface. Continuous records of potential-gradient have been accumulated for many years now at numerous stations widely distributed in geographical position. Observations are thus plentiful, and the changes in potential-gradient have been extensively studied, but why these changes occur is not yet completely understood.

At any one station the potential-gradient is subject to certain variations which are regular and well defined, and especially is this the case with the diurnal variation. But when the diurnal variations at different stations are compared, it is found that they do not all conform to one particular type, either in amplitude or in phase angle. Sometimes a difference even of period is found, the potential-gradient at certain stations having two maxima and minima each day, that at others having only one maximum and minimum.

Some degree of uniformity was introduced when Mauchly found[1] that the diurnal variation of potential-gradient over the oceans proceeds according to universal time and not local time. Shortly before this Hoffmann, comparing observations in polar regions, noticed[2] that the principal maxima and minima in the diurnal variation agreed if the hours were referred to universal time. Mauchly later analysed[3] the readings at observatories in different parts of the world and found indications of a diurnal oscillation in universal time. Whipple has shown[4] that this oscillation is similar to the diurnal frequency of thunderstorms all over the world.

Much earlier, Chree and Watson had pointed out[5] that the diurnal variation of potential-gradient at Kew Observatory is remarkably similar to that of atmospheric pollution. Now atmospheric pollution is evidently a local phenomenon, and

therefore it follows that potential-gradient is a function of local time, at any rate in polluted areas. More recently Whipple has shown[6] that the adoption of summer time in Great Britain, whereby the clock is put forward one hour in summer, has resulted in a displacement of the morning maximum and minimum of potential-gradient in summer, the G.m.t. of their occurrence now being one hour earlier than before the introduction of summer time. Thus the maximum and minimum in summer occur at the same hours as before by the civil clock, but are one hour earlier by the sun and by Greenwich time. More definite evidence that the diurnal variation of potential-gradient is affected by local phenomena could scarcely be expected.

The conclusion is irresistible that the diurnal variation of potential-gradient is governed by two factors, one of which is a function of universal time, the other of local time. At certain stations the universal term may predominate, at others the local term, and the magnitude of the local term obviously depends upon local conditions. For future interpretation of the diurnal variation of potential-gradient it becomes increasingly important to know what proportion of the local variation is due to the local term, and this presupposes a full knowledge of the dependence of the local term on local conditions. The purpose of the present paper is to investigate the effect upon the local electric field of the two types of atmospheric suspensoids: condensation nuclei, and the larger, opaque particles which at Kew Observatory are mainly particles of smoke.

The work of Chree and Watson shows the effect of the gross particles alone, though the unit employed by them was not the individual particle but the total weight of suspended matter. Special interest is attached to the correlation of the number of particles with electrical conditions, because the reaction is likely to depend upon the individual particles rather than on their total weight. Wait[7] and Builder[8] have investigated the effect of nuclei separately on potential-gradient and the electrical conductivity of the air. No work is known to the writer in which the effect of gross particles and condensation nuclei conjointly has been considered.

§ 2. OBSERVATIONS AT KEW OBSERVATORY*

At Kew Observatory observations of the electrical conductivity of the air have been made regularly for many years at 15 h. G.m.t. on fine days. Since the beginning of 1928 observations of the concentrations of gross particles and condensation nuclei have been made at the same time, as opportunity offers.

The observation of conductivity is made with a Wilson gold-leaf electrometer set on a tripod. Details of the method of observing are given in a memoir by Watson[9] in which reasons are adduced for believing that the element which is measured is the positive unipolar conductivity. The presence of the tripod results in a disturbance of the equipotential surfaces, the potential-gradient over the test plate of the electrometer being increased about six-fold. In winter, when the gradient is high, the congestion of the equipotential surfaces may produce a "saturation"

* Certain modifications of the routine observations have been made since the period dealt with in this paper.

effect the consequence of which is to give a value for the conductivity which is artificially low.

The observation of conductivity takes half an hour, and between the times at which the gold leaf is read observations of the concentration of condensation nuclei are made with an Aitken nucleus counter[10]. As nearly as possible at 15 h. a sample of air is introduced into an Owens jet dust counter[11] and the gross particles contained in this sample are subsequently counted* under a microscope. The observations of condensation nuclei and gross particles for the three years 1928–30 which are utilized in the present work are discussed in a paper by the writer[12].

The potential-gradient is recorded continuously at Kew Observatory by a Kelvin water-dropper. Particulars of the methods used to standardize the water-dropper are to be found in the Observatories Year Book[13]. The values of the potential-gradient and conductivity which form the subject of the present investigation are published in the same work.

It should be noted that the values of conductivity are the means for a period of about half an hour centred at 15 h., those of potential-gradient are the means for an hour centred at 15 h. The values of the concentration of condensation nuclei are the means for about twenty minutes centred at 15 h., while the values of the concentration of gross particles are instantaneous values. All the observations under discussion were made out of doors and in dry weather only.

In view of the reciprocal relation which is found to exist between the conductivity and each of the quantities (1) concentration of gross particles, (2) concentration of nuclei, (3) potential-gradient, it is more convenient to utilize the reciprocal of conductivity, which is here termed *resistivity*. The effect of saturation in winter is to give enhanced values of resistivity. It should therefore be borne in mind that when resistivity is high the true value is probably less than that which is given here.

Monthly means of the four elements under review are given in table 1. Since 15 h. is approximately the time of day at which a minimum of all four quantities is to be expected, the true monthly averages are considerably higher than those given in the table. A graphical representation of the results contained in table 1 is shown in figure 1. The resemblance of the behaviour of all four quantities is striking.

The annual inequalities may be resolved by harmonic analysis; the value of the element being

$$c_0 + c_1 \sin(t + \alpha_1) + c_2 \sin(2t + \alpha_2) + \dots,$$

where the unit of time is such that 1 year $= 2\pi$, the values of the amplitudes c_1 and phase angles α_1 of the first harmonic term are as given at the foot of table 1. The agreement in phase angle is noteworthy, the minimum of all four quantities occurring in the middle of July within a few days of each other. But no great significance can be attached to this coincidence, for most meteorological elements subject to a well-defined annual variation have turning points at about this time of year.

* According to an investigation carried out by Prof. Whytlaw-Gray, on behalf of the Atmospheric Pollution Research Committee of the Department of Scientific and Industrial Research, the number of particles given by the Owens instrument is about a fifth of the number present in the air and large enough to be seen by a microscope. The details of this investigation have not yet been published.

The association between these four quantities will now be examined in greater detail. The influence of particles and nuclei on resistivity will be dealt with first, and their effect on potential-gradient will be considered later.

Table 1. Monthly averages, Kew Observatory, 1928–30, 15 h. G.m.t. Dry weather.

Month	Gross particles (No. per cm^3)	Condensation nuclei (No. per cm^3)	Resistivity (10^{16} Ω.-cm.)	Potential-gradient (V./cm.)
January	1670	43,000	10·8	4·27
February	1560	35,000	11·2	4·61
March	880	31,000	6·9	3·42
April	414	23,000	3·2	2·75
May	231	26,000	3·1	2·21
June	60	15,000	2·4	1·81
July	60	21,000	2·0	1·54
August	86	15,000	1·9	1·67
September	226	19,000	2·5	2·26
October	345	27,000	3·8	2·76
November	1245	36,000	8·6	3·48
December	2400	55,000	11·8	4·50
Mean: c_0	765	29,000	5·7	2·94
Amplitude: c_1	964	14,300	5·0	1·49
Phase angle: α_1	82°	83°	78°	73°

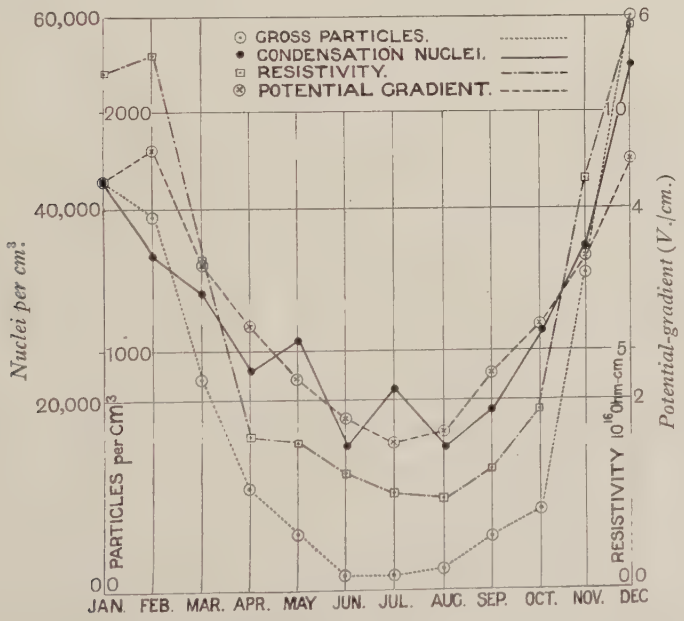


Figure 1. Annual variation, 1928–30, Kew Observatory, 15 h. G.m.t.

§3. SUSPENSOIDS AND RESISTIVITY

Individual observations have been grouped according to the concentration, firstly, of gross particles, and secondly, of condensation nuclei. For each such group the mean of the associated values of resistivity was determined. Since each group covers for the most part only a small range of numbers of particles or numbers of nuclei, as the case may be, the middle number of each group may be taken to represent with sufficient accuracy the average of the numbers of particles or of nuclei in the group. Thus we may associate the mean value of resistivity for each group with the middle number of particles or nuclei of that group.

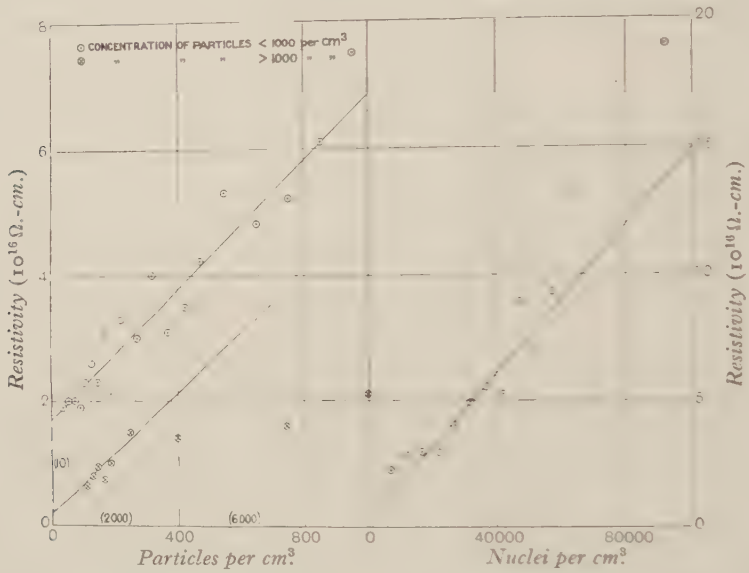


Figure 2. Values of resistivity for specified concentrations of gross particles and condensation nuclei.

The results thus obtained are represented graphically in figure 2. Owing to the extensive range in the concentration of gross particles found at Kew Observatory, the values of resistivity associated with these are represented on two scales. In the left-hand diagram of figure 2 circles refer to concentrations of particles less than 1000 per cm^3, crosses to concentrations greater than this figure. For the higher concentrations the scale is reduced tenfold, and the scale of resistivity is reduced in the same proportion. The two straight lines drawn in this diagram have the same equation, on their appropriate scales.

It may be seen that resistivity increases linearly with the number of gross particles until concentrations of the order of 3000 per cm^3 are reached. Beyond this point the resistivity increases more slowly with increasing pollution. The following argument shows that this is to be expected. The conductivity is composed of two parts, that due to small ions and that due to large ions. Ordinarily the contribution from

large ions is so small as to be negligible, but increasing pollution is associated with enhanced concentrations of large ions, and the contribution of these to the conductivity then becomes appreciable. By so much is the resistivity reduced.

It may be mentioned that a concentration of gross particles greater than 3000 per cm^3 at 15 h. is normally associated with conditions of mist or fog.

That the straight lines drawn through the points in the diagram do not pass through the origin is also to be expected, for if gross particles were completely absent the resistivity would not be zero.

The association between resistivity and the number of nuclei is shown in the right-hand diagram of figure 2. A straight line may be drawn to pass through the points with fair accuracy, though a curve may at first sight seem to give a better representation of their distribution. It is to be remembered, however, that high values of resistivity are probably enhanced artificially, and the departure from linearity indicated by these may not necessarily be valid. At the other end of the scale, where a departure from linearity appears to be suggested, there are theoretical reasons for believing that the curve should pass near the origin: the resistivity of air free from nuclei would depend only on the recombination of small ions, and the resistivity would then be such that on the scale used in the diagram it would appear to be very close to the origin*. All things considered, there seems to be no cogent reason at this stage for departing from a linear relationship.

That resistivity increases more rapidly with enhanced concentrations of nuclei is not in accordance with the argument put forward above, which applies equally to the case of nuclei, whereby a falling off in the rate of increase of resistivity with enhanced concentrations is to be expected. In spite of the fact that high values of resistivity are unduly augmented, we might have anticipated that the points representing these would lie below the optimum line, as was found in the case of gross particles. That the points representing low concentrations of nuclei should lie above this straight line is also unexplained. It is, however, almost certain that the points in the diagram in figure 2 may not be taken to represent purely the association between resistivity and the number of nuclei; for, if part of the resistivity is due to gross particles, that part is likely to be greater when nuclei are numerous than when nuclei are few, owing to the number of gross particles being greater when nuclei are more numerous.

It is necessary therefore to consider simultaneous values of all three quantities. The observations may be grouped according to values of resistivity, and mean values

* For air free from nuclei the resistivity

$$R = \frac{1}{ek}\sqrt{\left(\frac{\alpha}{q}\right)}\ \Omega.\text{-cm.},$$

where e is the electronic charge per ion, $15{\cdot}9 \times 10^{-20}$ coulombs;

k the mobility of small ions, 1·3 cm./sec. per V./cm.;

α the recombination coefficient of small ions, $1{\cdot}6 \times 10^{-6}$; and

q the rate of production of ions per cm^3 per sec.

If q is between 10 and 20, values appropriate to land stations, R is between $0{\cdot}1 \times 10^{16}$ and $0{\cdot}2 \times 10^{16}$ Ω.-cm.

of particles and nuclei determined for each group. The results obtained by grouping in this way are shown in table 2.

Table 2. Average concentrations.

The number of observations in each group is shown by the figures in brackets.

Resistivity (10^{16} Ω.-cm.)	Gross particles (No./cm^3)	Condensation nuclei (No./cm^3)	Resistivity (10^{16} Ω.-cm.)	Gross particles (No./cm^3)	Condensation nuclei (No./cm^3)
1·0–1·4 (12)	105	13,800	5·0–5·9 (11)	582	30,100
1·5–1·9 (11)	123	23,600	6, 7 (11)	1380	33,000
2·0–2·4 (10)	148	23,200	8, 9 (10)	1250	40,000
2·5–2·9 (16)	217	25,600	10, 11 (7)	2050	38,000
3·0–3·4 (12)	293	27,000	12, 13 (5)	4470	54,000
3·5–3·9 (9)	321	22,200	14, 17 (11)	5240	69,000
4·0–4·9 (10)	676	30,700	20–25 (4)	8050	71,000

A graphical representation may be made by selecting one of the variables as divisor and plotting the two resulting quotients against one another. Accordingly the middle value of resistivity for each group and the associated number of nuclei have been divided by the associated value of the number of particles, and the quotients are plotted in figure 3. Against each point in the diagram is placed the value of resistivity in units of 10^{16} Ω.-cm.

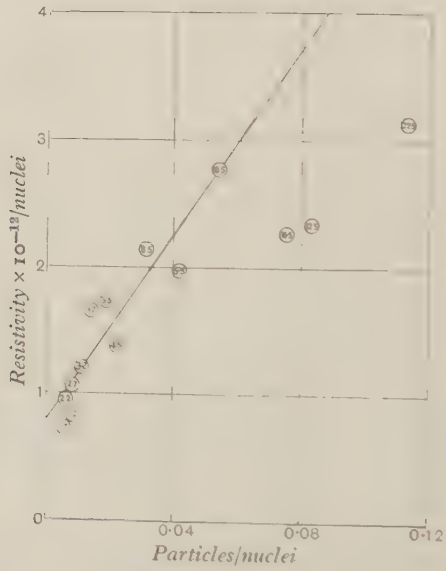


Figure 3. Illustrating the results contained in table 2.

Owing to the limitations of the observations of resistivity, no great significance can be attached to the points representing values of resistivity above 12×10^{16} Ω.-cm. It may be seen that the remainder lie fairly closely to a straight line.

The position of this straight line illustrates very clearly that the effect of gross particles on resistivity is not negligible. If the resistivity depended only on the number of nuclei we should expect the optimum line to be parallel with the horizontal axis. This is far from being the case.

The equation of the straight line to which figure 3 has been drawn is

$$10^{-12} R = 37P + 0{\cdot}77N \qquad \text{.....(1)},$$

where P and N denote the concentrations of particles and nuclei respectively per cm.^{3} P, N

§4. CORRELATION OF INDIVIDUAL OBSERVATIONS

So far we have been considering group means, and have found from these that resistivity appears to depend on both particles and nuclei more or less in accordance with a law represented by equation (1). It is necessary, however, to show that these results are borne out by the individual observations.

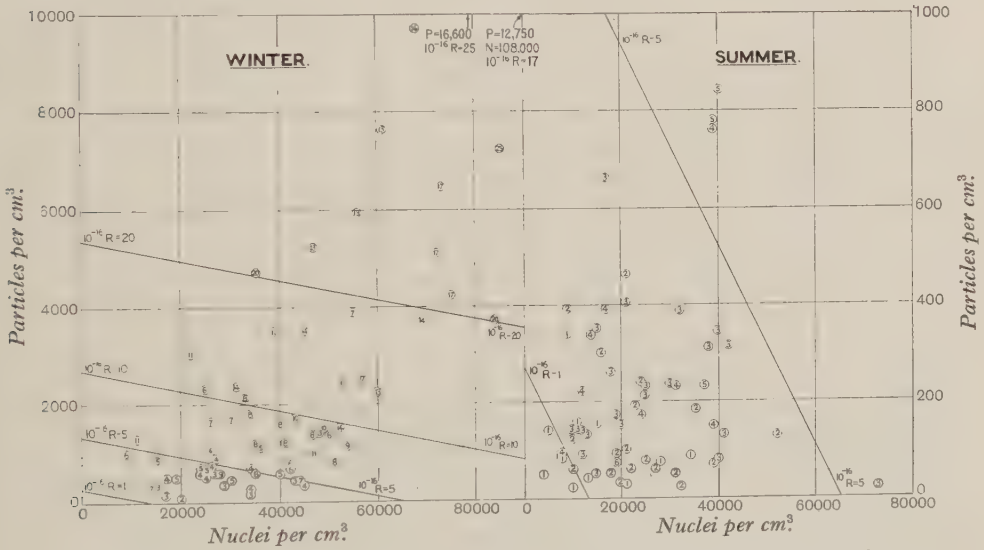


Figure 4. Simultaneous values of resistivity (10^{16} Ω.-cm.), and numbers of particles and nuclei per cm.^{3}

In figure 4 are represented the simultaneous concentrations of particles and nuclei at the time of resistivity observations. The associated value of resistivity, in units of 10^{16} Ω.-cm., is placed over the representative point. Owing to the extensive range in the number of gross particles the observations in winter and summer are represented in two separate diagrams; while the scale of nucleation is the same in each diagram, the scale of concentration of particles is ten times larger in the summer diagram than in the winter diagram. If a definite value of nucleation is selected, e.g. 30,000, it may be seen that while in summer only low values of resistivity are found on this ordinate, in winter considerably higher values are found on the same ordinate. This shows very clearly the effect on resistivity of the presence of gross

particles. Isopleths of resistivity have been drawn in the figure from equation (1). On the whole this equation fits the observations quite well.

A more detailed analysis may be made by a statistical correlation of the individual observations by the method of least squares. For this purpose all the available observations extending over the three years have been utilized, the number of simultaneous observations of resistivity and gross particles being 251, and the number of simultaneous observations of resistivity and nuclei being 240. It is desirable to treat separately the two seasons of summer (April to September) and winter (October to March) owing to the uncertainty of the high values of resistivity in winter.

The correlation coefficients are given in columns (1) and (2) of table 3, the pair

Table 3. Correlation coefficients.

	Resistivity					Potential-gradient							
	Total			Partial		Total			Partial				
	(1) r_{RP}	(2) r_{RN}	(3) r_{NP}	(4) $r_{RP.N}$	(5) $r_{RN.P}$	(6) r_{XR}	(7) r_{XP}	(8) r_{XN}	(9) $r_{XP.N}$	(10) $r_{XN.P}$	(11) $r_{XR.NP}$	(12) $r_{XP.RN}$	(13) $r_{XN.RP}$
Summer	·66	·42	·25	·63	·36	·72	·54	·39	·49	·31	·53	·12	·13
Winter	·66	·74	·71	·29	·52	·56	·51	·63	·12	·45	·14	·07	·34
Year	·74	·73	·69	·48	·46	·72	·60	·65	·28	·41	·40	·07	·32

	Means				Standard deviations			
	X	R	P	N	X	R	P	N
Summer	2·1	2·5	220	23,000	1·1	1·3	230	13,000
Winter	4·3	8·5	2100	40,000	1·7	5·8	2500	19,000
Year	3·2	5·2	1300	31,000	1·8	5·0	2000	18,000

of variables to which each coefficient refers being denoted by suffixes. The correlation coefficient between resistivity and the number of particles is 0·74 ± 0·03; that between resistivity and the number of nuclei is 0·73 ± 0·03. Corrections for annual variation may be made to these coefficients, using the harmonic expressions given at the foot of table 1. Corrected for annual variation, the coefficient between resistivity and particles is 0·73, and that between resistivity and nuclei is 0·70. These are practically the same as the uncorrected coefficients; thus the strong correlations are not due to similarity in the annual variation of the three variables.

A high correlation between resistivity and nuclei is to be expected: it is well known that the number of nuclei largely governs the number of small ions and hence the resistivity. That gross particles are also effective has not up to the present been well established, and the novelty of this feature renders it liable to suspicion. It may at first be thought that the strong correlation found between resistivity and the number of particles is due to the association between particles and nuclei; for both particles and nuclei are produced by combustion, and meteorological conditions which favour the accumulation of the one also favour the accumulation of the other. Hence there is a tendency for particles and nuclei to increase and decrease together. The correlation coefficients between these two from 169 observations are

given in the third column of table 3. It may be seen that in winter the correlation between the numbers of particles and nuclei is strong; in summer it is weaker but still positive.

It is necessary, therefore, to obtain a measure of the correlation between resistivity and the number of gross particles when the association of each with nuclei has been eliminated. This may be done by the method of partial correlation whereby the separate effects of each variable may be isolated. Partial correlation coefficients* are given in columns (4) and (5) of table 3, the notation being used in which the suffixes to the left of the point denote the two variables correlated and the suffix to the right denotes the variable the effect of which has been eliminated. It will be seen that in summer the partial coefficient between resistivity and the number of gross particles is greater than that between resistivity and the number of nuclei; in winter it is less. On the average, for the year as a whole, the partial coefficients are about the same. It follows that the high correlation between the number of particles and resistivity is not due to mutual associations with nuclei, and that gross particles can be at least as effective as nuclei in determining the resistivity.

Values of the means and standard deviations are included in table 3. The mean and standard deviation of R from simultaneous observations of R and P are not quite the same as the mean and standard deviation of R from simultaneous observations of R and N. Later it becomes necessary to correlate R with the potential-gradient, and the mean and standard deviation of R from this set of observations are also slightly different. These differences are however small, and the values given in the table are the means of all three. The same remark applies to the other variables.

If the regression equation is formed from the partial correlation coefficients and the values of the means and standard deviations are substituted, the equation expressing R in terms of P and N is found to be

$$10^{-16}R = 11 \times 10^{-4}P + 1{\cdot}2 \times 10^{-4}N + 0{\cdot}2 \pm 3{\cdot}0\,Ca \qquad \ldots\ldots(2),$$

where Ca denotes a casual number† the standard value of which is unity. If the casual term is ignored, the correlation coefficient between the two sides of this equation is 0·80.

The coefficients in equation (2) are not the same as those given in equation (1). This is because high values of R are included in (2), whereas in (1) these were deliberately ignored. In view of the influence of the high values of R, of which the validity is doubtful, on the coefficients in (2), and of the good fit of equation (1) to the individual observations as shown in figure 4, it seems likely that the coefficients in (1) may be more reliable than those in (2). The accuracy of the coefficients in either of these equations may not however be insisted upon.

It is to be noted that the residual term 0·2 in equation (2) is much less than the standard value of the casual term. A large residual term would mean that the

* The coefficients are determined by such formulae as

$$r_{12\cdot3} = (r_{12} - r_{13}r_{23})\,(1 - r_{13}^2)^{-\frac{1}{2}}\,(1 - r_{23}^2)^{-\frac{1}{2}}.$$

† This notation is due to Whipple[14]. In this particular example, 3·0 is the standard error of a determination of $10^{-16}R$ from the terms in P and N.

resistivity depended on other influences besides nuclei and particles. That the residual term is inappreciable suggests that the local resistivity may be wholly accounted for by the concentrations of nuclei and gross particles.

§ 5. RESISTIVITY AND POTENTIAL-GRADIENT

Before going on to examine the variations in potential-gradient with different concentrations of atmospheric suspensoids, it is desirable first to trace the association between potential-gradient and resistivity. The values of potential-gradient on occasions when the resistivity was within specified limits have been grouped together, and the mean for each group has been derived. These means are given in table 4, and the results are represented graphically in figure 5.

Table 4.

The number of observations in each group is shown by the figures in brackets.

Resistivity (10^{16} Ω.-cm.)	Potential-gradient (V./cm.)	Resistivity (10^{16} Ω.-cm.)	Potential-gradient (V./cm.)
0·9–1·1	1·56 (8)	4·1–4·5	3·46 (13)
1·2–1·4	1·50 (39)	4·6–5·5	3·55 (24)
1·5–1·7	1·59 (21)	5·6–6·5	3·93 (18)
1·8–2·0	1·83 (35)	6·6–7·5	4·87 (17)
2·1–2·3	1·86 (32)	7·6–8·5	4·72 (16)
2·4–2·6	2·12 (23)	9, 10	5·12 (12)
2·7–2·9	2·18 (25)	11–13	5·41 (14)
3·0–3·2	2·47 (14)	14–17	5·94 (17)
3·3–3·5	2·65 (18)	20–33	5·55 (11)
3·6–4·0	2·85 (29)	—	—

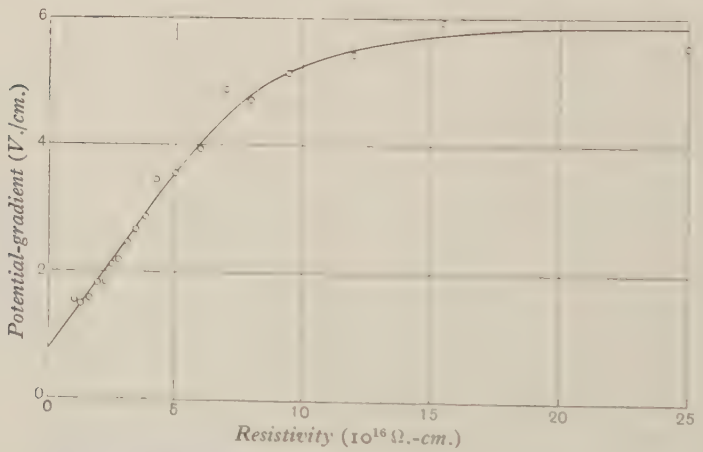


Figure 5. Values of potential-gradient for specified resistivity.

The points in the graph do not lie along a straight line through the origin as might have been expected from Ohm's law if the air-earth current at Kew had been constant. A straight line may be drawn through the origin so as to pass close to the

points representing values of resistivity less than about 4×10^{16} Ω.-cm., but the falling off for values of resistivity above this figure is very evident. The points representing high values of resistivity are undoubtedly less reliable, but the curvature is apparent for much lower values than can be disregarded.

A correlation of the individual observations has been made, all the available observations extending over the three years, 390 in number, being utilized. The correlation coefficients thus obtained are given in column (6) of table 3. From all the observations the correlation coefficient between potential-gradient and resistivity is $0{\cdot}72 \pm 0{\cdot}04$; corrected for annual variation by use of the harmonic expressions in table 1 the value is 0·57. The correlation is stronger in summer than in winter; thus changes in potential-gradient and resistivity are more closely related in summer than in winter. It is possible that the error attached to the high values of resistivity in winter reduces the coefficient for this season.

The coefficients are on the whole lower than might have been expected. They are of course worked out on the basis of linear regression. The graph in figure 5 suggests that the relation between resistivity and potential-gradient may not be linear. In order to allow for a departure from linearity, additional coefficients have been worked out, as a matter of interest, on the basis of a logarithmic regression; that is, potential-gradient was correlated with the logarithm of resistivity. The values of the coefficients were found to be 0·60 for summer, 0·65 for winter, and 0·79 for the year. Thus very little improvement in the value for the year is obtained by departing from a linear regression, and the value for summer becomes definitely reduced. The weakness in the correlation based on linear regression appears therefore to be due less to the non-linearity of the resistivity potential-gradient curve than to the occurrence of changes in potential-gradient in addition to those consequent upon changes in resistivity.

§6. SUSPENSOIDS AND POTENTIAL-GRADIENT

Simultaneous observations of suspensoids and potential-gradient may be analyzed by methods similar to those used to trace the association between suspensoids and resistivity. Accordingly the observations have been grouped according to the concentrations firstly of particles and secondly of nuclei, and mean values of potential-gradient have been derived for each group. The results thus obtained are illustrated in figure 6, two scales again being used in the case of gross particles owing to the difficulty of representing the extensive range of these on a single scale.

It may be seen from the diagram that potential-gradient increases with increasing concentration of particles, rapidly at first but more slowly later. The points do not appear to follow a straight line, the form of the graph resembling to some extent that between potential-gradient and resistivity. This suggests that gross particles affect potential-gradient only in so far as they lead to changes in resistivity. It will be seen later that this suggestion is supported by other considerations.

The points representing the association between potential-gradient and condensation nuclei are distributed closely about a straight line. It may be noted that this straight line does not pass through the origin.

If we consider simultaneous values of all three elements and group the observations according to the value of the potential-gradient, we may determine the mean numbers of particles and nuclei associated with values of potential-gradient within

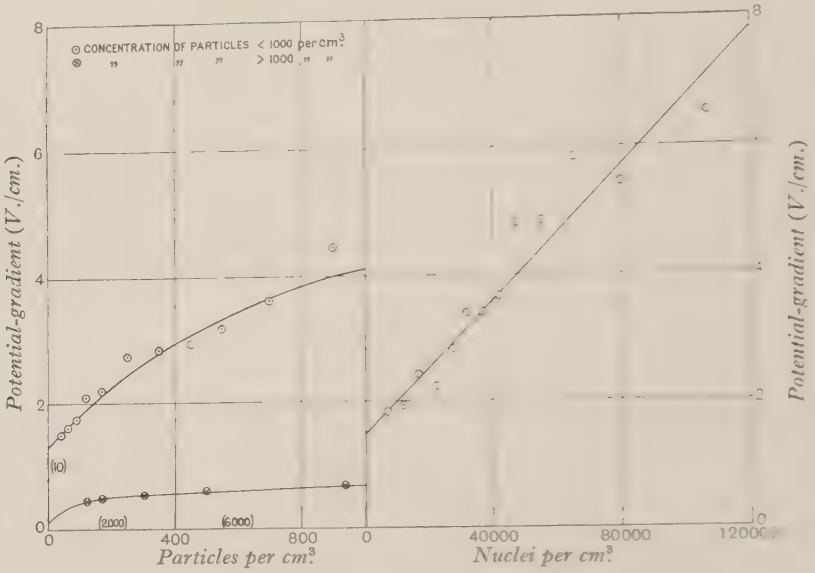


Figure 6. Values of potential-gradient for specified concentrations of gross particles and condensation nuclei.

specified limits. The results obtained by grouping in this way are shown in table 5. The last group in the table includes values of potential-gradient varying from 6·60 to 8·90 V./cm., the mean value for the group being 7·55 V./cm.

Table 5. Simultaneous concentrations.

The number of observations in each group is shown by the figures in brackets.

Potential-gradient (V./cm.)	Smoke particles (No./cm³)	Condensation nuclei (No./cm³)	Potential-gradient (V./cm.)	Smoke particles (No./cm³)	Condensation nuclei (No./cm³)
0·75–1·45 (20)	209	23,500	4·00–4·45 (6)	2310	45,500
1·50–1·95 (27)	271	20,000	4·50–4·95 (5)	1470	34,000
2·00–2·45 (24)	349	23,000	5·00–5·45 (10)	2100	44,000
2·50–2·95 (17)	877	29,500	5·50–5·95 (8)	2740	56,500
3·00–3·45 (13)	744	30,500	6·00–6·45 (5)	1700	42,000
3·50–3·95 (13)	568	31,000	> 6·50 (11)	4650	63,000

The results given in table 5 may be illustrated graphically by selecting one variable as a divisor and plotting the two resulting quotients. If the number of gross particles is selected as divisor the graph between the two resulting quotients is as shown in figure 7. Against each point in the diagram is placed the associated value of the potential-gradient. With but one exception the points are clustered fairly uniformly about a straight line. This straight line passes very close to, if not through,

the origin. If it does indeed pass through the origin the equation of the functional relation between X/P and N/P is reduced to the form

$$X = aN,$$

where X is the potential-gradient and N is the concentration of nuclei. This means that X is independent of P, the number of gross particles. X, N P

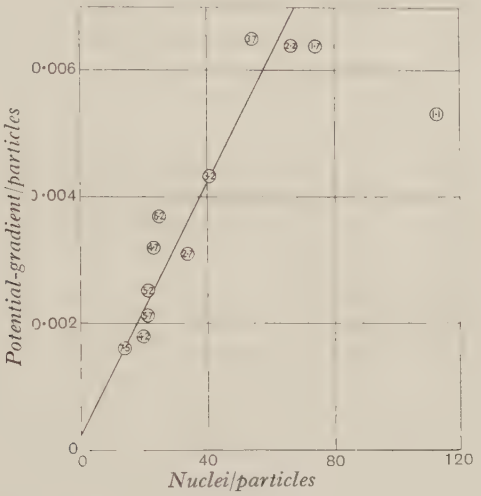


Figure 7. Illustrating the results contained in table 5.

§7. CORRELATION OF INDIVIDUAL OBSERVATIONS

A statistical analysis of the individual observations serves to show the strength of the association between suspensoids and potential-gradient, and gives some information as to why the association exists. Correlation coefficients between potential-gradient and the number of particles and nuclei are given in columns (7) and (8) of table 3, the two seasons of winter and summer being treated separately as in the case of resistivity. It may be mentioned that in working out the coefficients a few occasions of negative potential-gradient were rejected at the outset, but all the available positive values of potential-gradient have been utilized, the number of simultaneous observations of potential-gradient and gross particles being 286, and the number of simultaneous observations of potential-gradient and nuclei being 275.

The correlation coefficient between potential-gradient and the number of gross particles is 0·60 ± 0·04; corrected for annual variation by means of the harmonic expressions in table 1, the coefficient is 0·53. The coefficient between potential-gradient and the number of nuclei is 0·65 ± 0·03, and when corrected for annual variation 0·47. Neither coefficient is particularly high, though both are too significant to be neglected.

By combining these coefficients with those in column (3) between particles and nuclei, we may evaluate partial coefficients, representing the degree of association between potential-gradient and suspensoids when the mutual association between

the two types of suspensoids is eliminated. The partial coefficients thus obtained are shown in columns (9) and (10) of table 3. The partial coefficient between potential-gradient and the number of particles is only 0·28 and is practically negligible in winter, confirming the suggestion of figure 7 that the association between potential-gradient and the number of gross particles is but slight. The partial coefficient between potential-gradient and the number of nuclei is 0·41, a figure of some but not considerable significance.

The following question now arises. Is the association between potential-gradient and suspensoids due to a mutual association with resistivity, or does potential-gradient depend on suspensoids quite apart from, and in addition to, its variation with resistivity? This question may be expressed analytically as follows. The general linear equation between X, P, R and N may be written

a, b, c, d $$X = aR + bP + cN + d \qquad \text{.....(3).}$$

If b and c are zero, X is still a function of P and N in consequence of the relation

e, f $$R = eP + fN,$$

and the question is whether b and c are zero.

The coefficient of P in equation (3) is zero if the partial correlation coefficient of X with P when the association due to R and N has been eliminated is zero; and the same is true of the coefficient of N. Although partial coefficients connecting more than three variables are not ordinarily of much value, it seems worth while in this case to obtain the partial correlation coefficients between X and each one of R, P and N, when the effect of the other two has been eliminated. The coefficients thus obtained are as given in columns (11) to (13) of table 3.

It may be seen that the partial correlation coefficient between potential-gradient and the concentration of gross particles is so small as to be negligible, and this is true both in winter and in summer. Thus there is justification for stating that gross particles affect potential-gradient only in so far as they lead to changes in resistivity. The largest of the coefficients is that between potential-gradient and resistivity. This is however small in winter. As has been pointed out, the weakness of the correlation in winter may be due to errors in high values of resistivity. Condensation nuclei appear to have a small residual influence on potential-gradient in addition to their effect through resistivity, but this residual influence is practically negligible in summer. Consideration of the seasonal values of the coefficients suggests that changes in potential-gradient in summer may be due largely to changes in resistivity, while in winter there may be appreciable changes due to changes in nucleation.

If the residual influence of gross particles is neglected and a regression equation is formed between potential-gradient, resistivity, and nuclei,

$$\delta X = a\delta R + b\delta N,$$

the correlation coefficient between the two sides of this equation is 0·73 in summer, 0·65 in winter, and 0·74 for the year. These are not of considerable significance, and show that only about 50 per cent of the changes in potential-gradient may be

accounted for by local changes in the concentrations of atmospheric particles and nuclei.

Approximate values of the coefficients a and b in the equation may be determined. On substituting the values of the means and standard deviations we find

$$X = 0{\cdot}2 \times 10^{-16} R + 0{\cdot}3 \times 10^{-4} N + 1{\cdot}4 \pm 1{\cdot}2\,Ca \qquad \text{.....(4).}$$

The residual term is larger than the standard value of the casual term and may not therefore be neglected. It is of the same order as that of potential-gradient at sea at 15 h.; it represents the best available approximation to the potential-gradient which would have been found at Kew at 15 h. if the atmosphere had been free from pollution.

§ 8. THEORETICAL CONSIDERATIONS

The outstanding results of the statistical comparison are that: (1) The resistivity of the air at Kew depends on the concentration of gross particles and of condensation nuclei; and (2) the potential-gradient is related to resistivity and also to the concentration of nuclei. No direct influence of the gross particles on potential-gradient is apparent.

Resistivity. The usual way of regarding the association of resistivity with the amount of pollution is to say that the conductivity is due almost entirely to the small ions in the air. The number of small ions varies inversely as the number of condensation nuclei; for the rate at which small ions are produced is nearly constant, and the life of a small ion is terminated by collision with a nucleus, charged or uncharged. Thus the expectation of life of a small ion is inversely proportional to the number of nuclei, and it follows that resistivity is proportional to the number of nuclei. Moreover, since increasing pollution is associated with increased concentration of nuclei, resistivity is also likely to be proportional to the number of gross particles.

Conclusion (1) above is in general agreement with these ideas except that the statistics suggest that gross particles play an active part in determining the resistivity. The natural assumption to make is that some of the small ions collide with the gross particles and combine with them.

When nuclei are plentiful the equation of electrical balance between the number of small ions and the number of nuclei may be written[15]

$$q = 0{\cdot}61\eta Nn,$$

where q is the rate of production of small ions; q

η the recombination coefficient between small ions and large ions; η

N the total number of nuclei, charged and uncharged; N

and n the number of small ions. n

The factor 0·61 is adopted from the work of Nolan and O'Brolchain. It will be noticed that if the small ions had combined with equal readiness with large ions of opposite sign and with uncharged nuclei, and if the ions of either sign and the uncharged nuclei had been equally plentiful, this factor would have been $\frac{2}{3}$.

P β In order to allow for the possibility of collisions between small ions and gross particles we may add to this equation a term βPn, where P is the number of gross particles and β may be taken to represent a recombination coefficient between small ions and gross particles. Thus we write

$$q = 0{\cdot}61\eta Nn + \beta Pn \qquad \text{.....(5).}$$

Now the resistivity R is given by

$$R^{-1} = e\,(nk + N'/K),$$

k where k is the average mobility of the small ions;

N', K N' is the number of large ions and K their average mobility;

e and e is the electronic charge per ion, $15{\cdot}9 \times 10^{20}$ coulombs.

At Kew Observatory the average values of n and N' are of the order of 300 and 10,000 respectively. The value of k is about 1·3 cm. sec. per V. cm., that of K about 1/3000 cm./sec. per V. cm. Thus $N'K$ is small compared with nk, and we may write

$$R^{-1} = enk = 20{\cdot}7 \times 10^{-20} n \qquad \text{.....(6).}$$

Hence from (5) and (6)

$$10^{-18} R = 2{\cdot}9\eta\, N/q + 4{\cdot}8\beta P/q \qquad \text{.....(7).}$$

Nothing is known as to the value of q at Kew or as to its variations. Let us suppose that if q is not a constant the fluctuations in q are independent of those of N and P. For the present we are only concerned with averages, and we note that equation (7) is of the same type as equation (1) which was derived from the observations, viz.,

$$10^{-12} R = 0{\cdot}77 N + 37P \qquad \text{.....(1).}$$

Comparing coefficients and adopting the value of η found[15] at Dublin, $5{\cdot}7 \times 10^{-6}$, we deduce

$$q = 21 \text{ and } \beta = 17 \times 10^{-6}.$$

This value of q is improbably high. Perhaps we have taken too high an estimate for η.

For the ratio of the combination coefficients we are not dependent on the value of η or on the value of q at Kew. Equations (1) and (7) give

$$\beta/0{\cdot}61\eta = 37/0{\cdot}77 = 48.$$

This means that when nuclei and particles are equally numerous, the risk of a recombination between a small ion and a gross particle is about 50 times as great as the risk of a recombination between a small ion and a nucleus.

Now in estimating the risk of collision we may ignore the velocity of the larger particles, since they are very heavy compared with the small ions. Only the movements of small ions need be considered, and the risk of collision is therefore proportional to the area of the obstacle. Accepting the estimates 4×10^{-5} cm. and 4×10^{-6} cm. for the radii of particles[16] and nuclei[17] respectively, we find that the areas are in the ratio 100 : 1. That the recombination coefficients should be in the

ratio 50 : 1 is quite consistent, for it is possible that the proportion of combinations to collisions is lower with the large particles than with nuclei.

Potential-gradient. Potential-gradient may be regarded from two points of view, the statical and dynamical. According to electrostatics, if the charges in the atmosphere are distributed in uniform horizontal layers, then the potential-gradient at any point is proportional to the total charge above it. Hence we infer from the results of the statistical investigation that the total electric charge in the atmosphere increases when resistivity near the ground increases, and also, for constant resistivity, when the number of nuclei near the ground increases.

On the other hand, from the point of view of electrodynamics we note that the air-earth current I is numerically equal to the quotient of potential-gradient by resistivity. Hence from (4)

$$I = X/R = 0{\cdot}2 \times 10^{-16} + 0{\cdot}3 \times 10^{-4}\, N/R + 1{\cdot}4/R.$$

Thus the current decreases when the local resistivity increases, the number of nuclei remaining constant. This would be expected on the assumption that the voltage of the upper atmosphere is constant and the total resistance at all heights, $\int R\,dz$, increases with the local resistivity. The significance of the term in N/R in the last equation is however not at all obvious. But the details of the mechanism of the air-earth current, the parts played by eddy movements and by conductivity, are not known, and so the discussion of the observations under review cannot be carried further at present. More light will be thrown on the subject when observations of volume-charge are available at Kew.

§ 9. ACKNOWLEDGMENTS

I am indebted to Dr G. C. Simpson, C.B., F.R.S., Director of the Meteorological Office, for his interest and for permission to publish the work. I wish to express my best thanks to Dr F. J. W. Whipple, Sc.D., F.Inst.P., Superintendent of Kew Observatory, for his interest throughout the course of the investigation, and for his advice and generous help in the discussion of the results. My thanks are also due to Mr P. A. Sheppard, B.Sc., who has attended to the observational work since May 1929, and to Mr B. V. Bishop for assistance with the computations.

REFERENCES

(1) S. J. Mauchly. *Physic. Rev.* **18**, 161, 477 (1921); *Washington Res. Dep. Terr. Mag., Carnegie Inst.* **5**, 385 (1926).
(2) K. Hoffmann. *Beitr. Physik Atmos., Leipzig*, **11**, 1 (1923).
(3) S. J. Mauchly. *Terr. Mag.* **28**, 68 (1923).
(4) F. J. W. Whipple. *Q. J. R. Met. Soc.* **55**, 1 (1929).
(5) C. Chree and R. E. Watson. *Proc. R. S.* **105**, 311 (1924).
(6) F. J. W. Whipple. *Q. J. R. Met. Soc.* **55**, 351 (1929).
(7) G. R. Wait. *Terr. Mag.* **32**, 31 (1927).
(8) G. Builder. *Terr. Mag.* **34**, 281 (1929).
(9) R. E. Watson. *Geo. Mem.*, No. 45, London, 1929.

(10) J. Aitken. *Coll. Sci. Papers*, p. 236 (1923).
(11) J. S. Owens. *Advis. Comm. Atmos. Poll.*, p. 34 (1922).
(12) H. L. Wright. *Geo. Mem.*, No. 57, London, 1932.
(13) H.M. Stationery Office, London.
(14) F. J. W. Whipple. *Phil. Mag.* (i), 378 (1926).
(15) J. J. Nolan and C. O'Brolchain. *Proc. Roy. Irish Acad.* **38** A4, 40 (1929).
(16) H. L. Wright. *Loc. cit.*; see also *Advis. Comm. on Atmos. Poll.* pp. 36–8 (1923) for measurements of maximum diameters of particles under a microscope.
(17) Sir J. J. Thomson. *Conduction of Electricity through Gases*, 1, 189 (1928); J. J. Nolan, *Q. J. R. Met. Soc.* **55**, 173 (1929).

DISCUSSION

Dr F. J. W. Whipple. I am glad to have this opportunity to express my appreciation of the work done by the author in analysing the observations made at Kew Observatory of the numbers of particles in the air and of the conductivity and potential-gradient. It happens only too frequently that such observations, published or unpublished, accumulate without analysis.

One of the most striking developments in the paper is the demonstration that the resistivity of the air depends on the number of gross particles as well as on the number of nuclei. I mentioned this discovery of the author's some time ago and Prof. J. J. Nolan then declared that it was not in accordance with his observations in Ireland. Further comparisons are called for.

The explanation given by the author, that the gross particles are 100 times as effective as the nuclei in reducing the numbers of small ions by collision, may be correct, but we do not know that such collisions would destroy small ions. An alternative explanation is that the large ions associated with pollution are themselves bigger than those to be found in clean air. In fact there is no reason to suppose that the η of equation (5) is constant, and the factor to which the value 0·61 is assigned is probably variable.

The suggestion in figure 5 of the paper, that with resistivity zero the potential-gradient would be nearly 1 V. cm., invites comment. One can only surmise that the graph should bend sharply towards the origin. Unfortunately observations at Kew Observatory with air so pure that the resistivity falls below $10^{16}\,\Omega$.-cm. are too rare to provide good averages. The investigation will have to be extended to include observations in other parts of the country.

Prof. A. O. Rankine pointed out that according to the author the conductivity of the air appears to decrease as the number of nuclei increases—a surprising conclusion.

Author's reply. With regard to the apparent ineffectiveness of gross particles on the concentration of small ions in Ireland (reference 15), it is to be noted that variations in the function $q, 0{\cdot}61mN$ may not be expected to show any conspicuous dependence on P, but rather on the ratio P/N. This is evident from equation (5) which may be written in the form $q = 0{\cdot}61mN - \eta - \beta P, 0{\cdot}61N$. Dr Whipple's statement

that η is variable cannot, I think, be gainsaid, and to examine the extent of its variations is a problem for the future. Zero resistivity in air is difficult to imagine, but the resistivity at sea is about $0{\cdot}7 \times 10^{16}\,\Omega$.-cm., and the potential-gradient is of the order of 1 V./cm. These coordinates would lie on the graph drawn in figure 5, and so it would seem that if the graph eventually bends towards the origin it does so at a stage when lower resistivities than those normally experienced at sea are reached. That the graph drawn in figure 5 fails to pass through the origin is the geometrical illustration of the residual term in equation (4), though the magnitude of this term is not quite the same as that indicated in figure 5. It may be that we have to consider this as representing a universal term in the potential-gradient, uniform all over the world at a given instant.

The clue to the difficulty raised by Prof. Rankine is to be found in the second paragraph of § 8. I should perhaps have emphasized early in § 3 that the conductivity is preponderantly due to small ions owing to their extremely high mobility, and as a small ion is destroyed by collision with a nucleus, becoming in fact a large ion, it follows that the greater the number of nuclei the less the chance of survival of small ions, and hence the less the conductivity. Thus conductivity decreases with increasing concentration of nuclei.

621.352.12

THE TEMPERATURE COEFFICIENT OF THE SATURATED WESTON CELL

By P. VIGOUREUX and S. WATTS of the National Physical Laboratory

Communicated by Dr E. H. Rayner, October 17, 1932. *Read January* 20, 1933.

ABSTRACT. The effect of temperature on the electromotive force and on the performance of acid Weston cells has been studied from − 24 to + 40° C. The e.m.f. reached a maximum at about 3°, and at − 20° it was just over a millivolt less than the maximum. The cells behaved satisfactorily at − 16°, but at − 18° freezing took place gradually and was accompanied by a rapid decrease in e.m.f. At − 24° the electrolyte had completely solidified, and the e.m.f. was about 12 millivolts lower than the maximum. The frozen cells resumed their normal behaviour when kept at room temperature for about a day. Cooling increased the internal resistance of the cells. The e.m.f. temperature formula in common use does not hold below 0°, but a slightly different formula has been obtained, valid between − 20° and + 40°.

§ 1. PREVIOUS INVESTIGATIONS

The most extensive investigation of the temperature coefficient of Weston cells is due to Wolff*, who gave for the change of e.m.f. between 0° and 40° the formula

$$\Delta e = -40{\cdot}75\,(t-20) - 0{\cdot}944\,(t-20)^2 + 0{\cdot}098\,(t-20)^3$$

Δe, t in which Δe is the divergence in microvolts from the e.m.f. at 20°, and t is the temperature. It has subsequently been shown that acidity of the electrolyte does not affect the temperature coefficient to an appreciable extent†, and that the formula‡

$$\Delta e = -40{\cdot}6\,(t-20) - 0{\cdot}95\,(t-20)^2 + 0{\cdot}01\,(t-20)^3$$

expresses to a high order of accuracy the variation of electromotive force with temperature between the above-mentioned limits.

§ 2. DESCRIPTION OF CELLS EXAMINED

Twelve cells made in March 1932 were subjected to the temperature cycles. The cadmium amalgam constituting the negative pole contained only 6 per cent of cadmium in eight of the cells, but the usual 10 per cent in the other four. The reason for using a lower cadmium-content in some of the cells is that, for stability, both the solid and the liquid phases of the amalgam must be present together. At temperatures below − 10° C. the solid phase only of a 10-per-cent amalgam

* *Bull. Bur. Stand.* 5, 309 (1908).

† *Researches Electrotech. Lab.* No. 88 (1921).

‡ London Conference, 1908.

would be present, if extrapolation of experimental results from 0° downwards is valid*, whereas a 6-per-cent amalgam would secure stability only between − 35° and + 28°. Experiment subsequently showed that the 10-per-cent amalgam, contrary to expectations, behaved satisfactorily down to − 20°, but the 6-per-cent amalgam acted according to prediction, and gave abnormal values of e.m.f. at and above 30°.

In addition to the twelve cells subjected to the temperature test, six made in 1928 were connected permanently in parallel and used as comparison cells, and six made in 1929 were used individually as a check on the constancy of the comparison group. The solvent for the electrolyte of all the cells was sulphuric acid of decinormal strength.

Twelve cells, it might be thought, were hardly sufficient for the investigation, but the experience of many years acquired at the National Physical Laboratory, as well as the behaviour of the cells during the investigation, justified our views as to the stability and repetition of acid cells. At temperatures above 28° we could, it is true, use only the results obtained with the four cells containing 10-per-cent amalgam, but the exploration of that part of the range was only a verification of work often repeated, whereas the interest of the investigation lay chiefly with temperatures below 0°, at which the behaviour of standard cells had not as yet, as far as we are aware, been studied.

§ 3. EXPERIMENTAL DETAILS

A refrigerator, the tank of which contained a few hectolitres of a mixture of paraffin and machine oils kept constantly stirred, provided the required low temperatures. The twelve test cells were placed in a metal trough full of paraffin oil and provided with a heating resistance and a toluene thermostat. The trough could be lowered into the oil of the refrigerator tank. Stirring was effected by blowing dry air through small holes in a lead pipe resting on the bottom of the trough. By starting and stopping the refrigerator at intervals, it was easy to maintain in the trough a temperature constant to 0·05° at all points between − 24° and 5°. The comparison and reference cells were in an oil bath maintained at 20°. Mercury thermometers were used in both baths.

A Tinsley potentiometer giving single microvolts between successive studs of the last dial measured the difference in e.m.f. between each of the twelve test cells and the comparison group, and between each of the six reference cells and the comparison group. When the cells were all at 20° the sensitivity as read on the galvanometer was about 3 mm. for 10μV. At − 21° it was only about 0·8 mm., but even this was sufficient.

The time required to lower the temperature of the cells by 20° was about two hours. When the cells had reached the required temperature, half an hour was allowed to elapse and the eighteen measurements were made; about twenty minutes later, another set of readings was taken. It will be seen from table 1 that the two

* *Dict. of App. Phys.* **2**, 264.

Table 1. Measurements of the e.m.f. of twelve standard cells

Date, 1932	Summer time	Temperature of cells under test	6-per-cent amalgam			10-per-cent amalgam		
			e.m.f. of test cells minus e.m.f. of 1928 cells at 20° (μV.)	e.m.f. of test cells minus e.m.f. of test cells at 20° (μV.)	No. of cells used for the mean	e.m.f. of test cells minus e.m.f. of 1928 cells at 20° (μV.)	e.m.f. of test cells minus e.m.f. of test cells at 20° (μV.)	No. of cells used for the mean
April 26	—	20	+ 7	0	8	+ 7	0	4
May 3	—	− 4·1	+ 285	+ 278	8	—	—	—
	—	− 4·1	+ 279	+ 272	8	+ 284	+ 277	4
	—	− 10·2	+ 64	+ 57	8	+ 67	+ 60	4
	—	− 10·2	+ 62	+ 55	8	+ 64	+ 57	4
	16.30	− 15·35	− 249	− 256	8	− 246	− 253	4
	17.00	− 14·95	− 224	− 231	8	− 221	− 228	4
May 4	10.20	− 0·92	+ 341	+ 334	8	+ 344	+ 337	4
	10.50	− 0·95	+ 343	+ 336	8	+ 344	+ 337	4
	11.40	0·0	+ 354	+ 347	8	+ 354	+ 347	4
	12.20	0·0	+ 354	+ 347	8	+ 355	+ 348	4
	13.15	+ 1·0	+ 362	+ 355	8	+ 365	+ 358	4
	13.55	+ 1·0	+ 362	+ 355	8	+ 364	+ 357	4
	16.40	− 19·95	− 653	− 660	5	− 660	− 667	2
	17.15	− 19·9	− 652	− 659	5	− 663	− 670	2
May 5	10.45	− 1·75	+ 326	+ 319	8	+ 334	+ 327	4
	11.10	− 1·75	+ 325	+ 318	8	+ 332	+ 325	4
	11.45	0·0	+ 348	+ 341	8	+ 355	+ 348	4
	12.10	0·0	+ 348	+ 341	8	+ 355	+ 348	4
	12.50	+ 1·95	+ 363	+ 356	8	+ 368	+ 361	4
	13.40	+ 2·03	+ 364	+ 357	8	+ 370	+ 363	4
	14.15	+ 2·95	+ 369	+ 362	8	+ 381	+ 374	4
	14.50	+ 2·95	+ 367	+ 360	8	+ 373	+ 366	4
	15.25	+ 3·9	+ 365	+ 358	8	+ 372	+ 365	4
	15.55	+ 3·95	+ 366	+ 359	8	+ 371	+ 364	4
	16.30	+ 4·9	+ 362	+ 355	8	+ 368	+ 361	4
	17.00	+ 4·9	+ 361	+ 354	8	+ 369	+ 362	4
May 6	10.45	+ 20·0	+ 5	—	8	+ 6	—	4
	11.20	+ 20·0	+ 7	0	8	+ 7	0	4
	14.25	− 18·05	− 482	− 489	8	− 473	− 480	4
	15.35	− 18·0	− 483	− 490	8	− 475	− 482	4
	16.50	− 19·1	− 583	− 590	8	− 566	− 573	3
	17.30	− 19·3	− 607	− 614	7	− 605	− 612	3
May 7	10.50	− 19·0	− 579	− 586	3	− 581	− 588	2
	12.15	− 20·05	− 680	− 687	2	− 664	− 671	1
	12.45	− 20·05	− 671	− 678	1	− 668	− 675	1
May 9	11.10	− 7·65	+ 187	+ 180	8	+ 188	+ 181	4
	12.00	− 7·7	+ 183	+ 176	8	+ 190	+ 183	4
	16.15	− 21·0	− 796	− 803	1	The other 11 cells gave values between − 2000 and − 9000 μV.		
	17.00	− 21·15	e.m.f. was decreasing rapidly. Results varied between − 3000 and − 9000 μV.					
May 10	14.30	− 24·0	− 11000	—	8	− 12000	—	4
	15.10	− 24·0	− 11000	—	8	− 12000	—	4
	17.20	+ 20·0	− 99	—	8	− 52	—	4
	18.20	+ 20·0	− 84	—	8	− 50	—	4

Table 1. Measurements of the e.m.f. of twelve standard cells (*continued*)

Date, 1932	Summer time	Temperature of cells under test	6-per-cent amalgam			10-per-cent amalgam		
			e.m.f. of test cells minus e.m.f. of 1928 cells at 20° (μV.)	e.m.f. of test cells minus e.m.f. of test cells at 20° (μV.)	No. of cells used for the mean	e.m.f. of test cells minus e.m.f. of 1928 cells at 20° (μV.)	e.m.f. of test cells minus e.m.f. of test cells at 20° (μV.)	No. of cells used for the mean
May 11	10.00	+ 20·0	− 21	—	8	− 6	—	4
	10.45	+ 20·02	− 22	—	8	− 6	—	4
	15.15	− 18·1	− 526	− 506	6	− 500	− 494	4
	16.05	− 18·1	− 526	− 506	6	− 506	− 492	3
May 12	10.30	− 18·1	− 519	− 499	5	− 505	− 491	2
	12.30	+ 20·0	− 20	—	8	− 14	—	4
	14.00	+ 20·0	− 17	—	8	− 13	—	4
	16.30	− 16·1	− 352	− 335	8	− 339	− 326	4
May 13	11.00	− 16·1	− 335	− 335	8	− 324	− 330	4
	12.30	− 16·15	− 338	− 338	8	− 329	− 335	4
	14.30	− 16·15	− 337	− 337	8	− 326	− 332	4
May 14	13.00	+ 20·0	+ 1	—	8	+ 6	—	4
May 17	17.00	20·0	− 2	—	8	+ 9	—	4
May 19	16.30	20·0	− 1	—	8	+ 9	—	4
May 26	—	20·0	+ 4	—	8	+ 10	—	4
May 30	—	20·0	+ 6	—	8	+ 10	—	4
June 9	—	20·0	+ 6	—	8	+ 9	—	4
June 18	—	20·0	+ 6	—	8	+ 7	—	4
July 8	13.45	20·0	+ 6	—	8	+ 8	—	4
	14.55	20·0	+ 5	—	8	+ 9	—	4
	16.40	26·0	− 264	− 271	8	− 263	− 272	4
	17.50	26·0	− 266	− 273	8	− 264	− 273	4
July 9	10.50	20·0	+ 8	—	8	+ 8	—	4
	12.00	20·0	+ 8	—	8	+ 10	—	4
July 11	11.00	30·0	− 492	Abnormal	—	− 474	− 482	4
	12.00	30·0	—	,,	—	− 479	− 487	4
	13.45	34·0	—	,,	—	− 717	− 722	4
	14.30	34·0	—	,,	—	− 717	− 722	4
	16.00	36·97	—	,,	—	− 910	− 913	4
	16.45	37·0	—	,,	—	− 906	− 909	4
July 12	11.20	20·0	+ 5	—	8	+ 2	—	4
	12.30	20·0	+ 7	—	8	+ 6	—	4
	14.55	40·0	—	Abnormal	—	− 1116	− 1120	4
	15.50	40·0	—	,,	—	− 1113	− 1118	4
	16.45	40·0	—	,,	—	− 1109	− 1114	4
July 13	11.15	20·0	+ 5	—	8	+ 5	—	4
	12.30	20·0	+ 5	—	8	+ 5	—	4
July 14	10.45	20·0	+ 6	—	8	+ 6	—	4
	11.35	20·0	+ 6	—	8	+ 5	—	4
	12.35	22·0	− 78	− 84	8	− 78	− 83	4
	14.30	22·0	− 77	− 83	8	− 78	− 83	4
July 15	10.30	17·41	+ 103	+ 96	8	+ 104	+ 96	4
	12.00	17·54	+ 100	+ 93	8	+ 100	+ 92	4
	13.45	20·0	+ 8	—	8	+ 8	—	4
	14.45	20·0	+ 7	—	8	+ 8	—	4

sets were always in excellent agreement, indicating that the lag was very small. It was consequently possible to complete the low-temperature investigation in ten days, during which 736 individual measurements were made; they were spread over sixteen different temperatures between − 24° and 5°. In addition, the test cells were often raised to 20°, and measured at that temperature. A similar procedure was observed for measurements at room temperature and above.

§ 4. BEHAVIOUR OF THE CELLS

The summary of results in table 2 shows that the electromotive force reaches a maximum at about 3°; the curve connecting electromotive force and temperature resembles a parabola with vertex upwards and axis perpendicular to the temperature axis as in figure 1. Down to − 16° the electromotive force reached the

Table 2. Relation between e.m.f. and temperature

Temperature (° C.)	Experiment*		Formula − experiment (μV.)	
	Actual points	Best curve	Formula in general use	Formula proposed
− 20	− 672	− 674	139	− 3
− 19	− 582	− 577	124	+ 2
− 18	− 487	− 485	108	+ 3
− 16	− 324	− 313	76	0
− 15	− 230	− 233	62	− 2
− 10	68	75	18	− 6
− 7·7	179	174	9	+ 3
− 4·1	276	276	11	+ 5
− 1·75	322	322	10	+ 5
− 1	335	333	7	+ 5
0	346	346	6	+ 4
1	356	355	5	+ 2
2	359	361	3	+ 1
3	365	363	3	0
4	362	362	4	0
5	358	358	4	− 1
10	—	301	0	− 5
14	—	205	2	− 3
17·5	92	92	3	0
20	0	0	0	0
22	− 83	− 83	− 2	+ 2
26	− 272	− 272	− 4	+ 5
30	− 484	− 484	− 7	+ 5
34	− 722	− 722	− 5	+ 6
37	− 912	− 912	− 3	0
40	− 1118	− 1118	+ 6	0

* Excess in microvolts of the e.m.f. over the e.m.f. at 20°.

value expected from the curve in less than an hour, and suffered no subsequent change. When the cells were cooled to − 18°, the e.m.f. also reached the expected value within an hour, after which it would remain constant for several hours; but then one of the twelve cells would suffer a gradual decrease in e.m.f. Then another cell, and then a third, would behave in a similar manner. It was found on examination that the electrolyte of the cells undergoing the abnormal decrease

was gradually becoming solidified. The process was slow at − 18°, for after twenty hours at that temperature five out of the twelve cells were partly frozen, but it is probable that the electrolyte in all of them would have become solidified had the experiment been continued. A similar test was tried at − 16°, but the cells behaved normally during its twenty-four hours' duration. We conclude that Weston cells, or at least those made with solvent of decinormal acidity, are satisfactory at − 16°.

The process of solidification is much faster when the temperature is lowered to − 19° or − 20° instead of to − 18° only; but even at − 21° one of the cells behaved normally for about half an hour. In evaluating the mean electromotive

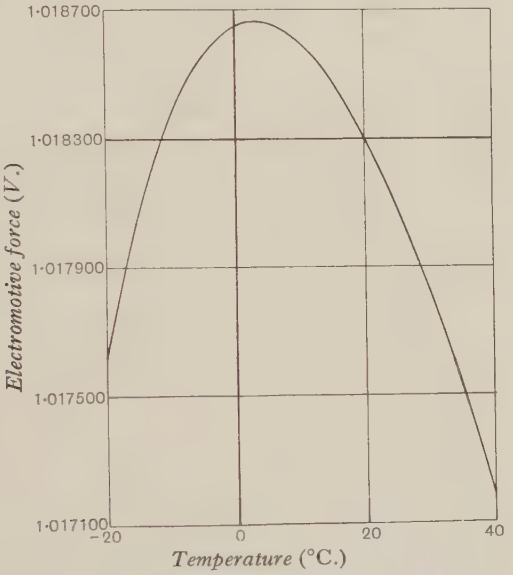


Figure 1. Relation between the electromotive force of the Weston (saturated) cell and its temperature.

force at any temperature, only those cells were considered whose behaviour was normal; above − 18° results on all the cells were used, but below this temperature the cells which exhibited a continuous decrease in e.m.f. were, of course, not taken into account.

After the freezing to − 24°, an undoubtedly severe treatment, the bath was warmed, and in an hour the temperature had reached + 20°. Half an hour later the 6-per-cent amalgam cells still gave about 100 μV., and the 10-per-cent about 50 μV., below the normal value. This, and the subsequent behaviour, are shown in table 1. Recovery was complete in about a week.

Want of space precludes us from giving here the individual results from which table 1 was constructed, but it should be mentioned that the behaviour of the cells was not so uniform that they all had the same value at all temperatures; nevertheless the maximum and mean deviations from the mean, even at − 18°, did not exceed 22 and 11 μV. respectively, so that the mean of the twelve cells may be regarded as a fair average, of accuracy 5 μV.

§ 5. DERIVATION OF A FORMULA

The temperatures in table 1 are not all in round numbers, but it is easy to obtain the first two columns of table 2 by plotting a graph and applying the small corrections required. The figures in the second column are the means of the values of the 6-per-cent and 10-per-cent amalgam cells for all temperatures below 30°, and of the latter only for higher temperatures. A curve drawn on a large scale yielded the figures in column 3, which agree closely with the experimental points. The formula in general use to-day, viz.,

$$\Delta e = -40{\cdot}6\,(t-20) - 0{\cdot}95\,(t-20)^2 + 0{\cdot}01\,(t-20)^3,$$

does not agree with experiment below 0°, as shown by column 4, which gives the differences between that formula and the experimental curve. A cubic formula seems inadequate to represent the actual state of things, but by going one step further it is possible to obtain for the whole range the same accuracy which the cubic formula gives for the positive half. From the figures in the last column, which are differences between the formula

$$\Delta e = -39{\cdot}39\,(t-20) - 0{\cdot}903\,(t-20)^2 + 0{\cdot}0066\,(t-20)^3 - 1{\cdot}5\times 10^{-4}\,(t-20)^4$$

and the experimental curve, it will be seen that only in one case is the discrepancy as large as 6 μV. The formula can be simplified by referring the temperatures to 3°, thus

$$\Delta e = 363 - 1{\cdot}5\,(t-3)^2 + 1{\cdot}68\times 10^{-2}\,(t-3)^3 - 1{\cdot}5\times 10^{-4}\,(t-3)^4.$$

Other formulae have been tried with maxima at 3·2, 3·3 and 3·4°, but they do not give as good agreement as the above, and it seems that the maximum e.m.f. occurs very close indeed to 3°, which might, for some purposes, prove a convenient temperature for the standardization and comparison of cells.

§ 6. DISCUSSION OF RESULTS

The temperature of maximum electromotive force determined by this investigation is in exact agreement with the results of Wolff*. Although it is possible that the variations of electromotive force with temperature depend on other factors besides the concentration of the electrolyte, the negligible temperature coefficient of unsaturated cells, indicating as it does that the influence of such factors is very small, suggests that the minimum solubility of cadmium sulphate, whether in water or decinormal sulphuric acid, occurs at a temperature of 3°. This result is in agreement with the solubility determinations of Mylius and Funk† and of von Steinwehr‡, but at variance with those of Cohen and Wolters§, who assign − 9° to the point of minimum solubility.

* *Loc. cit.*

‡ *Z. phys. Chem.* **88**, 229 (1914).

† *Ber. deuts. chem. Ges.* **30**, 824 (1897).

§ *Z. phys. Chem.* **96**, 253 (1920).

The phenomena observed at low temperatures can be explained from the freezing-point and solubility data* for cadmium sulphate. The solubility curve of $3CdSO_4.8H_2O$ meets the freezing-point curve at $-17°$, when a solid of composition $CdSO_4 + 14{\cdot}5H_2O$ is precipitated. The cell can, therefore, be expected to function satisfactorily above this temperature, and probably, with suitable amalgams, up to $74{\cdot}5°$, at which temperature another transformation takes place. The slow rate of solidification at temperatures above $-19°$ might be caused by the formation of the metastable $CdSO_4.7H_2O$, which would lower the temperature of solidification of the solution, before it gradually suffered transformation into the stable salt $3CdSO_4.8H_2O$.

§ 7. INTERNAL RESISTANCE OF CELLS

An approximation to the internal resistance was obtained by observing the sensitivity of the potentiometer at various temperatures. The calibration was effected by replacing the cells by a resistance box. The figures below, representative of the values obtained, are only a rough approximation. The small resistance of the six comparison cells has been allowed for where necessary.

Temperature (° C.)	20	5	0	−10	−20	−24 (completely frozen)
Average internal resistance (Ω.)	800	850	1700	2500	5000	40,000

The average resistance of the reference cells made in 1929 seems to lie between 500 and 600 Ω.

§ 8. ACKNOWLEDGMENTS

We wish to thank the Superintendent of the Physics Department of the National Physical Laboratory, Dr Kaye, who gave us facilities for carrying out the low temperature work in his department; the several members of his staff for their help, and Dr E. H. Rayner, Superintendent of the Electricity Department, who instigated the research, for his encouragement throughout the investigation.

* Mylius and Funk, *loc. cit.*

538.71

A NEW APPARATUS FOR THE MEASUREMENT OF THE EARTH'S MAGNETIC FIELD

By L. F. BATES, Ph.D., Reader in Physics, University College, London

Received October 18, 1932. *Read December* 16, 1932.

ABSTRACT. A new method for the determination of the horizontal and vertical components of the earth's magnetic field is described. A small cylinder of mu-metal is wound with fine wire whose ends are connected to wires forming a torsional suspension. This cylinder is placed with its centre at the mid-point of a Helmholtz coil system whose axis is adjusted to coincide with the direction of the component of the earth's field which is to be measured, the suspension being perpendicular to the component. The current through the coil system can be adjusted so that there is no deflection of the suspended cylinder when the current in the solenoid wound upon it is reversed, in which case the calculated field produced by the coils is equal to the required component. The sources of error peculiar to the method are examined and their elimination is discussed. It is shown that measurements of high accuracy are possible.

§ 1. INTRODUCTION

For a complete knowledge of the earth's magnetic field in a particular region it is necessary to determine the magnitudes of the horizontal and vertical components of that field. The horizontal component may be accurately found by several methods, the most rapid being that of Schuster and F. E. Smith*, of which a simple modification was recently described by the writer†. The vertical component is usually calculated from a knowledge of the horizontal component and the angle of dip, the latter being generally determined with an earth inductor in series with a galvanometer.

F. E. Smith‡ has outlined a direct method for the measurement of the vertical component. A small coil is rotated with constant frequency about a horizontal diameter in the magnetic meridian, the ends of the coil being connected to a vibration galvanometer. The vertical field acting on the rotating coil may be annulled by passing a current through a Helmholtz galvanometer coil system, whose axis is vertical and whose mid-point coincides with the mid-point of the rotating coil. When no current flows through the galvanometer, the current in the coil system is exactly that required to neutralize the vertical component of the earth's field, so that this component can be calculated from the dimensions of the coil system and the magnitude of the current.

* F. E. Smith, *Phil. Trans. R. S.* A **223**, 175 (1922).
† L. F. Bates, *J. Sci. Inst.* **8**, 324 (1931).
‡ F. E. Smith, *loc. cit.* 177.

The writer was not aware that any other methods had been reported until his work with the apparatus described below had been completed and the following account—which has been left entirely unaltered—had been written. His attention was then drawn to a method described and tested with customary precision by the late Dr D. W. Dye*.

In Dye's method a very light coil is suspended so that it may rotate about a horizontal axis in the magnetic meridian, the plane of the coil being in the meridian. When supplied with alternating current of the resonant frequency, the coil vibrates in a manner similar to a coil-type vibration galvanometer under the influence of the vertical component of the earth's field. The mid-point of the coil is situated at the mid-point of a Helmholtz coil system whose axis is vertical.

A current supplied to the Helmholtz coil system is adjusted until the resonance vibrations disappear, when the vertical component of the earth's field is exactly neutralized in the region occupied by the vibrating coil. The vertical component is calculated from a knowledge of the current and the dimensions of the coil system. Using the refinements possible at the National Physical Laboratory, Dye thus measured the vertical component with an accuracy of one part in twenty thousand.

The apparatus now to be described enables both the horizontal and the vertical components to be rapidly determined with considerable accuracy. With the present construction the mean error of a determination is of the order of one part in a thousand, but with further refinements greater accuracy is obtainable.

§ 2. THEORY OF THE METHOD

The method of determining the vertical component, for which the apparatus was primarily designed, will be first described. In figure 1*a*, XX represents a small

Figure 1*a*.

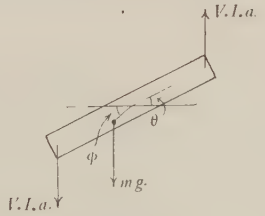


Figure 1*b*.

cylinder of mu-metal, chosen because of its very high permeability in low fields, on which is wound a solenoid of thin insulated wire. The ends of the wire are brought to B and C, where they are soldered to the ends of two fine phosphor-bronze galvanometer suspensions, AB and CD. The whole system is then mounted so that AD lies horizontally and in the magnetic meridian, whilst the axis of XX is adjusted to be horizontal and perpendicular to the line AD.

The passage of a current of a few milliamperes through the solenoid is sufficient to make the mu-metal, initially unmagnetized, a strong magnet. The vertical component of the earth's field acts upon this magnet, which tends to set with its

* D. W. Dye, *Proc. R. S.* A 117, 434 (1927).

magnetic axis vertical, and, consequently, the cylinder is deflected. The deflection is reversed by reversing the current through AD, and no deflection occurs if the vertical component of the earth's field is annulled.

The cylinder XX is mounted with its centre at the mid-point of a Helmholtz coil system whose axis is vertical, as shown in figure 2. A steady current is supplied to the coils, and by repeated adjustment of the current the vertical component of the earth's field may be entirely compensated in the region occupied by XX. When this condition is satisfied no deflection of XX occurs on reversing the current in AD. If exact compensation is not attained, the direction of the deflection of XX on reversal of the current in the solenoid will depend upon whether the compensating current is too large or too small to annul precisely the vertical component.

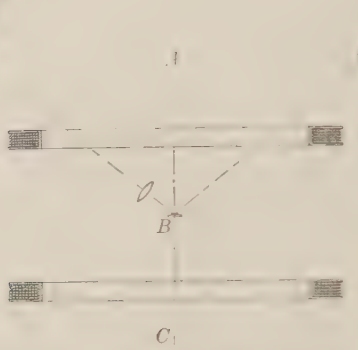


Figure 2. To measure the horizontal component: planes of coils are vertical with meridian on line ABC (i.e. the figure is plan view). To measure the vertical component: planes of coils are horizontal, with meridian perpendicular to plane of paper at B (i.e. the figure is an elevation). The magnet suspension is at B and perpendicular to plane of paper.

It is therefore possible to determine when exact compensation is attained, and hence, from a knowledge of the compensation current and the geometric constants of the Helmholtz coils, to calculate the magnitude of the vertical component of the earth's field. The several factors upon which the sensitiveness of the apparatus depends will now be described.

I, a Let I be the intensity of magnetization of the mu-metal cylinder, a its area of
l, m cross-section and l its length. Let the mass of the cylinder and solenoid be m, and
x suppose that the centre of gravity is at a distance x from the line about which the
θ cylinder turns. Let the axis of the cylinder make an angle θ with the horizontal, and let the angle between the line from the centre of gravity perpendicular to the
ϕ axis of rotation and the horizontal be ϕ, when the magnetization is such that the vertical component of the earth's field gives rise to forces in the directions shown by the arrows in figure 1*b*.

Then, for equilibrium,

$$VIal \cos\theta + mgx \cos\phi = T\gamma \qquad \text{.....(1)},$$

where V is the vertical component of the earth's field, T is the torsion constant of the system and γ is the angle through which it is twisted. The forces need not, of course, be coplanar. On reversing the magnetizing current the magnetic forces shown in figure 1*b* are reversed, the angles θ, ϕ and γ are decreased by $\delta\theta$, and we now have

V, T γ $\delta\theta$

$$-VIal\cos(\theta-\delta\theta)+mgx\cos(\phi-\delta\theta)=T(\gamma-\delta\theta) \quad \ldots\ldots(2).$$

Subtracting (2) from (1), we have

$$VIal\{\cos\theta+\cos(\theta-\delta\theta)\}+mgx\{\cos\phi-\cos(\phi-\delta\theta)\}=T\delta\theta,$$

or, since $\delta\theta$ is small,

$$VIal\{2\cos\theta+\delta\theta\sin\theta\}-mgx\,\delta\theta\sin\phi=T\delta\theta,$$

whence

$$\delta\theta=2VIal\cos\theta/(T+mgx\sin\phi-VIal\sin\theta) \quad \ldots\ldots(3).$$

Under the chosen experimental conditions the axis of the cylinder is practically horizontal, i.e. $\theta=0$. Also, when the vertical component of the earth's field is almost compensated, the field V must be replaced by a small field, δV, and we have

δV

$$\delta\theta=2\delta VIal/(T+mgx\sin\phi) \quad \ldots\ldots(4).$$

The second term in the denominator of equation (4) may be made vanishingly small by decreasing x by the method described below, so that, in the limit, we may write

$$\delta\theta=2\delta VIal/T.$$

Hence the sensitiveness of the apparatus may be increased by increasing I—which means the use of solenoids of many turns—by increasing the dimensions of the cylinder, and, finally, by decreasing the torsion constant.

The value of I is limited by the magnitude of the current which may be passed through the solenoid together with the high demagnetization factor of the cylinder. The latter determines to some extent the dimensions of the cylinder, which are also limited by the mechanical strength of the suspension and by the fact that, unless the cylinder is small, it does not lie within the region over which the field produced by the Helmholtz coils can be regarded as constant. In fact, given that the position of the centre of gravity of the system has been properly adjusted, the torsion constant of the suspension is the factor which mainly governs the sensitiveness of the apparatus.

§ 3. EXPERIMENTAL DETAILS

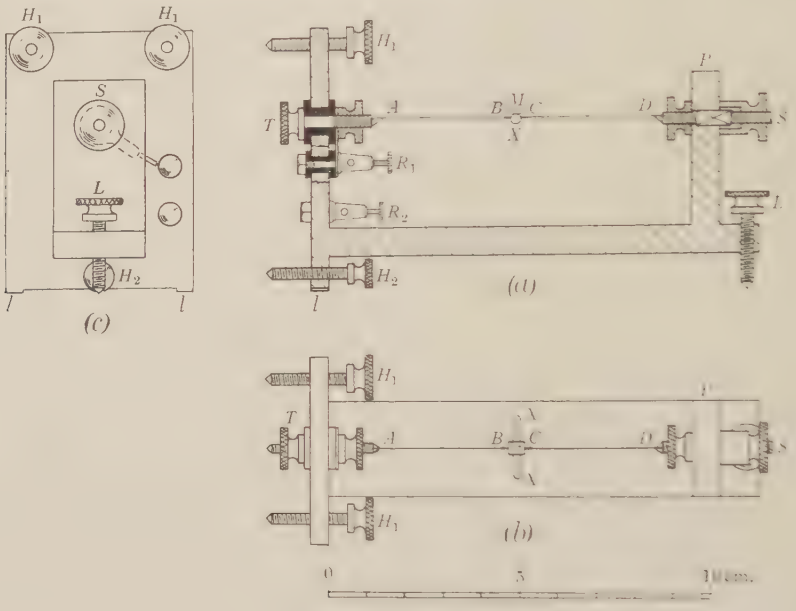
Figures 3*a*, 3*b* and 3*c* represent respectively the elevation and plan views of the actual apparatus as arranged for the determination of the vertical component. The frame was constructed of non-magnetic brass. From figures 3*a* and 3*b*, it will be seen that the tension of the suspension could be adjusted by means of the screw S, which passes through a square hole in the brass pillar P.

The axis of the cylinder XX could be brought into a horizontal position by means of the torsion head T. The latter passed, as shown, through an ebonite bush,

and it was in electrical communication with the insulated terminal, R_1. A second terminal, R_2, directly screwed to the brass support, allowed a current to be sent through the suspended system.

The electrical arrangements are shown in figure 4. A shunted millivoltmeter was used for measuring the current supplied to the Helmholtz coils. The latter were each of 30 turns of mean diameter 40 cm. Their centres were fixed (for convenience) 17·2 cm. apart in all experiments. The millivoltmeter was frequently calibrated and its shunt often changed so that the current determinations were free of bias.

The levelling screw L together with two projections l l, from the left-hand side of the base, figures 3*a* and 3*c*, enabled the suspension AD to be arranged as nearly



Figures 3*a*, 3*b*, 3*c*.

horizontal as the slight sag, which was inevitably present, permitted. The dimensions of the frame were chosen with the view of making this sag as small as possible without greatly diminishing the sensitiveness.

In the diagrams a simple device for clamping the suspended system, when in transit, is not shown. Two pointers, attached to the base of the instrument vertically below the line AD, which were particularly helpful in fixing the position of the apparatus with respect to the magnetic meridian, are likewise not shown. A cardboard case provided with mica windows was used to shield the suspended system from the effects of air currents.

No electromagnetic or frictional damping devices were necessary to bring the suspended system to rest, as with a little practice it was easy to determine the

equilibrium position when small oscillations persisted, and large oscillations were easily reduced by proper manipulation of the reversing switch shown in figure 4.

The cylinder XX was 1·5 cm. long and 1·4 mm. in diameter. These dimensions satisfied the conditions mentioned in the discussion of the theory of the method. Several such cylinders were wound and mounted, and experiments were made with those referred to as nos. 1, 2, 3 and 4 respectively. The solenoids on nos. 1 and 2 were wound by hand, the ends of the wire being looped as in figure 5, and then drawn taut. No. 1 was wound with 40 turns of no 36 s.w.g. copper wire, and no. 2 with 80 turns of no. 42 s.w.g. eureka wire. The wire was double-silk-covered and was passed through molten paraffin before being wound, in order to ensure more perfect insulation.

The solenoids on cylinders nos. 3 and 4 were wound on a lathe. A small hole was bored through the middle of each cylinder, and the solenoid was wound in two layers. The winding was started at the middle of the cylinder. Half of the

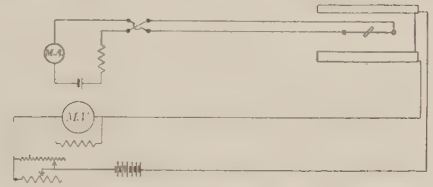


Figure 4.

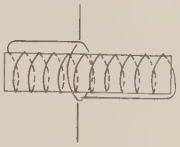


Figure 5.

cylinder to one side of the hole was completely wound and then the other half, the free end of the wire being passed through the hole. Special precautions were necessary to prevent short-circuiting of the solenoid where the wires passed through the hole. Cylinder no. 3 was wound with 160 turns of no. 42 s.w.g. eureka wire, and cylinder no. 4 with 140 turns of no. 36 s.w.g. copper wire.

On mounting the suspension it was generally found that the centre of gravity of the system was too far below the line AD. The centre of gravity was therefore raised by attaching small quantities of molten picein or of sealing wax to each end of the upper surface of the solenoid, until the period of torsional oscillation about AD became large. This was best done by loosening the screw S, lowering the system on to a support, and adding or removing wax with a hot copper wire. Cylinders nos. 3 and 4 were very symmetrically mounted and only required the addition of very small quantities of wax.

The phosphor-bronze galvanometer strip suspension which gave satisfactory results was 0·0025 × 0·0005 in. in cross-section, the lengths AB and CD being approximately those shown in figure 3. The mirror M was mounted with picein on the upper side of the cylinder, before adjustment of the centre of gravity. It was made from a small piece of microscope cover glass coated with platinum.

The optical arrangements for recording the rotation of the cylinder are shown in figure 2. The scale was usually conveniently placed about 50 cm. from the mirror M; this distance could, of course, be increased. A suspension was considered satisfactory if the reversal of a current of not more than 10 mA. in the solenoid

produced a deflection of about 10 cm. on the scale when the vertical component of the earth's field was not compensated.

§ 4. SOURCES OF ERROR

The first possible source of error to be discussed is the effect of the current in AD upon the poles of the cylinder. For the purpose of calculation, we may look upon AD of figure 3*a* as a long straight wire carrying a current i. The pole strength of the mu-metal cylinder is Ia, and a force approximately equal to $(2i\ \frac{1}{2}l)\ Ia$ will act on each pole. Hence the system experiences a couple of electromagnetic origin equal to $4iIa$. Since i and I always have the same sign, the sense of this couple is constant.

The effect of the current in AD will therefore be to cause a constant displacement of the cylinder, but it cannot affect the determination of the vertical component unless I varies in magnitude with the direction of the current in the solenoid. In the experiments described below, the maximum value of i was 30 mA., and l was 1·5 cm. Hence the maximum value of the couple was 0·008 Ial.

Thus, even if the numerical value of I changes by as much as 1 per cent on reversal of the current in the solenoid, the deflection $\delta\theta$ will only be proportional to 0·00008 Ial. Since it is easily possible to measure the compensating current to one part in 800, the limiting deflection due to uncompensated field is proportional to $V/800\ Ial$ or 0·0005 Ial. The error due to the lack of constancy of I can only become important when measurements are made to an accuracy of greater than one part in 10,000.

In any case, with a comparatively high field in the solenoid it is difficult to believe that the value of I can change by as much as 1 per cent on reversal. If, however, it is proved that the change is appreciable, then the error can be greatly reduced by running a wire parallel to and just below AD, so that the solenoid current passes along AD and returns through the lower wire, whose magnetic field will almost neutralize that of AD near the poles of the cylinder.

Another source of error arises when the line AD does not coincide with a horizontal line in the magnetic meridian passing through the centre of the cylinder. A slight displacement of AD from the meridian can easily be shown to give rise to very little error. A slight displacement from the horizontal, however, gives rise to serious error. For let AD make a small angle α with the horizontal. Then the horizontal component of the earth's field may be resolved into components, $H\cos\alpha$ and $H\sin\alpha$, respectively parallel and perpendicular to AD. The latter component will produce a couple approximately given by $H\sin\alpha\,.\,Ial$. If α is of the order of 1°, this couple is approximately 0·004 Ial, i.e. about eight times greater than the smallest measurable current difference. There is, fortunately, every reason why α should be very much less than 1° with an apparatus properly constructed and set up.

In any case the error may be avoided by taking a further value of the compensating current with the whole apparatus turned through 180° in a horizontal

plane, when the direction of the couple due to the horizontal component of the earth's field will be reversed, while the vertical couple remains unchanged. This rotation is much facilitated by the provision of the pointers mentioned in an earlier section, care being taken that the centre of the cylinder is always placed in the same position with respect to the Helmholtz coils.

The most serious source of error arises if the axis of the coils is not strictly vertical. If measurements correct to 1 part in 1000 are required, a slight departure from the vertical will not be serious, but for more accurate measurements it is clear that the coil system must be wound on an accurately turned cylinder which can be adjusted with its axis vertical.

§ 5. EXAMINATION OF ERRORS

The exact current necessary for the compensation of the vertical field may be found in two ways, either by successive adjustment until no deflection occurs on reversal of the solenoid current, or by interpolation from the deflections produced on reversal when the current in the coils is a little too small or a little too large. The first method was found to be much the easier and the more satisfactory, although, in the tests made, both gave identical results within the limits of experimental error. The first method was therefore always employed.

In the following experiments the procedure usually adopted was for one observer to record the deflection produced on reversal of the solenoid current. A second observer adjusted the current in the coils until the first observer signalled that exact compensation was attained. The compensating current was then read. Two observers are not necessary, but in these tests they made for more speedy progress.

To investigate the presence of errors of the type described in the first portion of the last section, as well as any due to the heating of the suspension, the magnitude of the solenoid current was varied. The tests were confined to currents which produced a fairly high degree of magnetization of the mu-metal. For cylinder no. 1 the successive values obtained for the compensating current with a current of 12 mA. through the solenoid were 79·7, 79·9, 79·9 and 79·65, giving a mean value of 79·79 divisions on the shunted millivoltmeter. With a current of 30 mA. through the solenoid, and other experimental arrangements unchanged, the successive values were 79·8, 79·9, 79·8 and 79·7, giving a mean value of 79·80 divisions. There was, then, no change with the higher current, and since the corresponding scale deflections were larger with the higher current, it was usually employed with cylinders wound like no. 1. Heavier currents caused appreciable sagging of the suspension owing to expansion, and produced frequent changes of zero due to convection currents; they were therefore not employed.

The errors due to the presence of the horizontal component of the earth's field were next investigated; these should be very small if the cylinder is properly mounted. With the system on which the last tests were made, and a current of 30 mA. through the solenoid, the following values of the compensating current

were obtained: 79·5, 79·4 and 79·5 divisions. The apparatus was then turned through 180° in a horizontal plane and the following readings were obtained: 79·6, 79·65, 79·6, 79·75 and 79·55 divisions. It was then restored to its original position, and the following readings were recorded: 79·4, 79·6 and 79·55 divisions. The mean values for the two positions were therefore 79·49 and 79·63 respectively, so that the final error due to the horizontal component of the earth's field was certainly small. The value for the vertical component of the earth's field obtained from these determinations was 0·3931 gauss, a value consistent with the laboratory structure.

Tests were also made with cylinder no. 3. The sensitiveness of the apparatus with this cylinder was very high, a deflection of about 30 cm. being recorded on the reversal of a current of 12 mA. in the solenoid when no current flowed in the Helmholtz coils. It was now not necessary to use more than 12 mA., owing to the large number of windings in the solenoid.

With the apparatus in one suitable position the following values for the compensating current were obtained: 78·40, 78·35, 78·40, 78·30 and 78·30 divisions. The apparatus was then turned through 180° in a horizontal plane and the values now obtained were 78·50, 78·50, 78·55, 78·50, 78·50, 78·50, 78·35, 78·55 and 78·50 divisions. It was then replaced in its former position and the values obtained were 78·45, 78·45, 78·55, 78·55 and 78·55 divisions. The mean of all these values gave a value of 0·3926 gauss for the vertical component of the earth's field.

The first set of these readings gives a lower mean value than the last set. It should be recorded, however, that in the first set the zero error of the millivoltmeter was much larger than ever observed in any other set, probably due to a stray thermoelectric effect in the circuit. The first set may therefore be in error. If the mean of the second and third sets only is taken, the value of the vertical component obtained is 0·3929 gauss, which compares well with the earlier value of 0·3931 obtained with the first system. The three sets of readings show, however, that the cylinder system was well mounted.

In order to determine the magnitude of the error likely to arise through the line *AD*, about which the cylinder turns, failing to coincide with a horizontal line in the magnetic meridian, the apparatus was purposely inclined at various angles to the horizontal.

With cylinder no. 2, the levelling screw *L*, figure 3, was first screwed into the base to the maximum extent, so that the line *AD* made an angle of 1° 12′ with the horizontal. With the apparatus so placed that the screw *S* was to the north of the cylinder, the neutralization current was found to be 81·27 divisions. The apparatus was then turned through 180° when the neutralization current was 79·89. On restoring the apparatus to the first position the neutralization current was 81·31 divisions. The mean value for the two positions was therefore 80·59 divisions.

The suspension was then made as horizontal as possible, by sighting, and the currents with the screw *S* north and south of the cylinder were now respectively 79·88 and 81·25 divisions, the mean value being 80·57. It will be observed that the neutralization current for the first position had become the smaller.

The screw *L* was then unscrewed to the maximum extent, so that the line *AD*

made an angle of 5° with the horizontal. The two neutralization currents were now 77·82 and 83·53 divisions respectively, giving a mean of 80·68. Hence, for a displacement of over 6° a difference of only one part in 800 between the mean values was produced.

This error may be made insignificant, for it is clear that for some setting of the apparatus the neutralization currents must be identical. In that event the line *AD* will be strictly horizontal if it lies in the magnetic meridian, which is practically the case.

In determining this setting a simple graphical method was found very effective. The screw *L* was screwed in as much as possible and a datum line was marked upon its head. The neutralization current was measured with the apparatus so placed that *L* was to the north of the cylinder. The screw was then turned through one complete turn and the neutralization current was again measured. This procedure was repeated several times. The apparatus was then turned through 180° and the measurements were repeated with *L* to the south of the cylinder. A set of results, obtained with cylinder no. 2, are plotted in figure 6. It will be seen that the

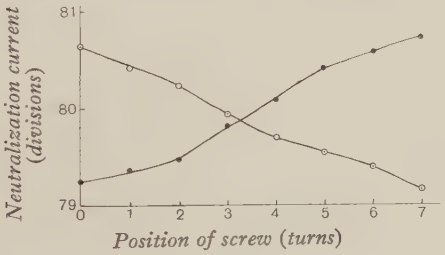


Figure 6. Cylinder no. 2.

best position of the apparatus was that obtained by turning the screw *L* through 3·2 turns from the zero position. This was done, and the values of the two currents were then found to be 80·10 and 80·18 divisions respectively, giving a mean of 80·14. The values of the vertical component of the earth's field calculated from the mean of the observations, from the graph of figure 6, and from the measurements with the best setting were respectively 0·3927, 0·3925 and 0·3928 gauss.

Experiments on these lines were also made with cylinder no. 4, the current in the solenoid being 24 mA. Determinations of the neutralization currents were made for six recorded positions of the screw *L*. The value of the vertical component found from the mean of these six determinations was 0·3928, that found graphically was 0·3931, and, finally, that obtained from determinations when the apparatus was set in the most favourable position given by the graph was 0·3931 gauss. The corresponding graph was very similar to that of figure 6.

§ 6. COLLECTED RESULTS

The results of the measurements with the various cylinders are given in the table.

Specimen no.	Date of determination	V	Remarks
1	11. v. 32	0·3931	Zero error large
3	18. v. 32	0·3926	
2	27. v. 32	0·3939	
2	30. v. 32	0·3927	Direct mean
		0·3925	Value from graph
		0·3928	Best position
4	1. vi. 32	0·3928	Direct mean
		0·3931	Value from graph
		0·3931	Best position

The last two sets of results are considered to be the most accurate. However, even the earlier results show that it is possible, without taking any rigorous precautions, to determine the magnitude of the vertical component to within 1 part in 400 in a few minutes. If higher accuracy is required, the procedure described in the preceding section may be used, and then the accuracy appears to be limited only by the accuracy of the current-measuring arrangements, the surface of the table on which the brass frame rests, and the accuracy of the winding of the Helmholtz coils.

It should be recorded that the above experiments indicated that, provided care was taken to ensure adequate insulation of the wires drawn through the hole, the drilled cylinders such as nos. 3 and 4 were the most satisfactory. Moreover, the cylinders wound with no. 36 s.w.g. wire were much the easier to adjust and mount symmetrically, and they exhibited far fewer displacements of zero, which probably arise from slight changes in the positions of the thin ends of the solenoid.

There appears to be no reason why the apparatus, including the coil system, should not be designed in a portable form suitable for work in the field.

§ 7. DETERMINATION OF THE HORIZONTAL COMPONENT

It is clear from the diagrams of figure 3 that the apparatus may also be used for the determination of the horizontal component of the earth's magnetic field. This is important, for a value of this component can be conveniently obtained by an independent method, and hence the behaviour of the apparatus can be checked. From figure 3 it will be seen that it is necessary to turn the whole apparatus through 90° in a vertical plane. The Helmholtz coils are now, of course, set with their axis parallel to the magnetic meridian. The axis about which the cylinder rotates is made vertical by means of the levelling screws H_1 and H_2.

The sources of error associated with the measurement of the horizontal component are essentially those described in the earlier portion of this paper. The chief error is due to the effect of the vertical component of the earth's field on the

mu-metal cylinder when the axis about which the latter rotates is not strictly vertical. A slight displacement of this axis from the meridian is relatively unimportant, but its displacement from the vertical in the meridian gives rise to serious error. In the latter case, if the axis of rotation makes a small angle α with the vertical, the vertical component of the earth's field can be resolved into components, $V \cos \alpha$ and $V \sin \alpha$, respectively parallel and perpendicular to the axis of rotation. The latter component will produce a couple approximately equal to $V \sin \alpha . Ial.$ If α is of the order of 1°, this couple is approximately equal to 0·009 *Ial.*

This source of error may be eliminated by determining the neutralization currents with the apparatus in two positions differing by 180° in a horizontal plane, the mean value of these two currents being equal to the correct neutralization current. The most favourable position for the apparatus may be found by the method outlined in an earlier section.

For this, it is expedient to place the apparatus with the axis of the cylinder at right angles to the magnetic meridian and parallel to the line joining the screws H_1. The centre of the cylinder coincides with the centre of the coil system. The suspension is rendered as vertical as possible by adjustment of the screws H_1 and H_2, the suspension being sighted against two plumb-lines during the last operation; if the solenoid is well constructed this means that the axis of rotation is practically vertical.

Let us suppose that in this position the mirror faces north. A datum mark is made on the head of the screw H_2, which may then be turned through a series of known angles and the corresponding neutralization currents may be determined. The apparatus is now turned through 180° in a horizontal plane, so that the mirror faces south, and the neutralization currents determined for the same series of positions of the screw H_2. A set of such determinations is reproduced in figure 7*a*, cylinder no. 4 being used in this case because of its symmetry; the current in the solenoid was 24 mA. We see that the best position of the apparatus is obtained by turning the screw through two complete turns from the zero position of the table. This was actually the position in which the apparatus was initially placed by sighting with the aid of the plumb-lines. The correct neutralization current obtained from the graph was 35·90, and from the mean of the observations, 35·89 divisions. The value found from a series of observations specially taken with the apparatus in the best position was 35·90 divisions. The mean value of the horizontal component thus found was 0·1755 gauss.

Figure 7*a* effectively illustrates the appreciable difference between the two values of the neutralization current for the two 180° positions of the apparatus, when the axis of rotation is tilted in a north-south plane. On theoretical grounds the effect of an inclination in the east-west plane should be much less. Such inclinations were produced by turning the head of one of the screws H_1 through known angles; a set of observations so obtained with cylinder no. 4 is reproduced in figure 7*b*. The mean value of the neutralization current obtained from the graph of figure 7*b* is 35·93 and from the mean of the observations 35·98 divisions; the agreement is satisfactory. It will of course be recognized that the table on

which the apparatus shown in figure 3 is placed must be levelled as accurately as possible.

The above determinations were checked by an entirely independent method. The apparatus shown in figure 3 was removed, and between the coils was placed a magnet and mirror system suspended by a very fine silk fibre. The neutralization current was now determined by a method described elsewhere*. The dimensions

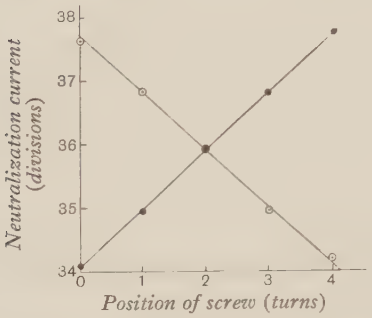


Figure 7*a*. Cylinder no. 4, north-south displacement.

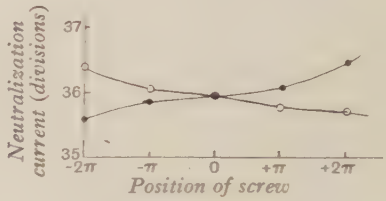


Figure 7*b*. Cylinder no. 4, east-west displacement.

of the magnet were very nearly the same as those of the mu-metal cylinder, so that the neutralization conditions should have been the same in both experiments, except that for the check determination the coils had to be rotated slightly. The corrected mean value obtained for the neutralization current for several displacements of the coils was 35·94 divisions; the check is satisfactory.

If the axis of the Helmholtz coil system is not accurately horizontal and in the magnetic meridian, the experimental value of the neutralization current will be too high. Hence in an arrangement designed for very accurate determinations, both the coil system and the apparatus of figure 3 should be mounted on turntables. The coil table could then be placed in known positions and the position giving a minimum value of the neutralization current determined by a simple graphical method.

§ 8. ACKNOWLEDGMENTS

I again have very much pleasure in thanking Prof. E. N. da C. Andrade for his keen interest and support. I also wish to record my thanks to Mr A. Maxwell Lewis for his assistance in many series of test measurements.

* L. F. Bates, *loc. cit.*

DISCUSSION

Mr R. S. WHIPPLE. Has the author considered the application of an instrument of this type to geophysical surveying? A higher degree of accuracy than the figures given in the paper would be necessary, but not improbably, with care in construction, this could be obtained. It would then be possible to use the same instrument in the field for the determination of the horizontal and vertical forces. Members might be interested to know that a Dye magnetometer has recently been constructed in which the Helmholtz coils are wound on metal formers, instead of on a marble cylinder as in the original instrument. In this way a portable instrument of high accuracy has been produced.

Mr GUILD suggested that, as a null method was employed, the magnetizing solenoid need not be part of the suspended system. With a fixed solenoid, a small piece of mu-metal could be used in the suspension.

AUTHOR'S reply. In reply to Mr Whipple: I see no reason why the instrument should not be made with greater sensitivity than I have so far employed, in a portable form suitable for geophysical surveying; in fact, I think it could be made very compact indeed.

In connection with Mr Guild's suggestion, I would like to point out that the difficulties of construction would be very considerably increased in providing fixed solenoids to magnetize the mu-metal, and I think additional sources of error would arise. In any case, the clamping of the moving system would be much more difficult.

547·532 + 547·551·1]: 536.63

A METHOD FOR THE DETERMINATION OF THE SPECIFIC HEATS OF LIQUIDS, AND A DETERMINATION OF THE SPECIFIC HEATS OF ANILINE AND BENZENE OVER THE APPROXIMATE RANGE 20° C. TO 50° C.

By ALLAN FERGUSON, M.A., D.Sc.

AND

J. T. MILLER, B.Sc., East London College

Received November 18, 1932. *Read December* 16, 1932.

ABSTRACT. The specific heats of aniline and benzene are determined by measuring the electrical power E^2/R necessary to hold the temperature of a calorimeter and its contents steady at various temperature-excesses θ. A cooling curve is then taken in order to evaluate the constant K in the equation

$$(MS + W)\, K\theta^{\frac{5}{4}} = E^2/RJ,$$

where S is the specific heat of the liquid at a temperature θ above that of the surroundings.

The specific heat of aniline at any temperature T between 20° C. and 50° C. is given by

$$S = \cdot 495_0 + \cdot 000275\,(T - 20) + \cdot 0000035\,(T - 20)^2,$$

and that of benzene by

$$S = \cdot 395_8 + \cdot 00125\,(T - 20).$$

§ 1. INTRODUCTION

There exists an increasing need for fairly precise values of certain thermal constants of organic liquids and solutions, which play an important part in technological processes—we refer especially to specific heats, specific volumes and latent heats. In this communication we confine ourselves to a measurement of specific heat and its temperature-variation over a moderate range.

A glance at the relevant entries in the standard books of tables shows that the figures quoted in them must be of very doubtful value*, so that it becomes necessary to consult the actual papers, and to assess for oneself the reliability of the final figures. We may say here that, in comparing our figures with those of other experimenters, we have retained only the results of Messrs Williams and Daniels† for benzene, which were obtained by the use of an adiabatic calorimeter, and of Dr H. R. Lang‡, who used the continuous flow method, for aniline.

We believe that there is room, and need, for an experimental method which

* For instance, Landolt and Börnstein give for the specific heat of aniline, 0·510, 0·490, 0·491 at 20° C. and for the specific heat of benzene, 0·4502, 0·441, and 0·438 at 50° C.

† *J. Amer. Chem. Soc.* **46**, 903 (1924).

‡ *Proc. R. S.* A **118**, 138 (1928).

can be used in precision work and yet, in a simple form, give with such ordinary apparatus as is to be found in every reasonably equipped works laboratory, results that can be depended upon to 1 per cent or better.

The method to be described does, we think, approximate to these requirements, and when it is used under the conditions we indicate the results may be depended upon, as far as the physical work is concerned, to at least 1 part in 500, which was the standard we set ourselves at the outset.

Experience of the method has convinced us that, with standardization of corresponding accuracy, we may obtain results accurate to 1 part in 1000. Our results are certainly reproducible to about that order with the present apparatus.

§ 2. METHOD

In brief, the method consists in supplying by means of a heating coil, and measuring, the power necessary to hold a copper calorimeter and its contents at temperatures, for example, 5° C., 10° C., 15° C. ... higher than that of the surrounding medium, which is usually kept at about room-temperature. After the calorimeter and its contents have been raised to a temperature of about 50° C. the power is switched off, and a cooling curve is taken. This completes the experimental observations.

Then at any temperature $\theta°$ above that of the surroundings, if the power required to hold the temperature stationary is E^2/R watts, we have — E, R

$$-\frac{dQ}{dt} = -(MS + W)\frac{d\theta}{dt} = \frac{E^2}{JR} \qquad \text{.....(1),}$$

where J has the value 4·186 or 4·1845, according as a watt is defined as 10^7 ergs per second or in terms of the international ampère. In our case the appropriate value is 4·1845, although to the order of accuracy contemplated in these experiments the difference between the two values is without numerical significance.

$-dQ/dt$ in the above equation represents the rate of loss of heat from the calorimeter and contents, M the mass of liquid contained in the calorimeter whose water-equivalent is W, and S the specific heat of the liquid at the given temperature. — Q, t; M; W, S

If, now, we can show that

$$-\frac{d\theta}{dt} = K\theta^n,$$

and can determine the values of K and n, we have — K, n

$$(MS + W)K\theta^n = E^2/RJ \qquad \text{.....(2)}$$

as the equation which gives the specific heat S at a temperature $\theta°$ above that of the surroundings*. For each value of E^2/R we can work out a value of S at the corresponding temperature. The unit in which S is measured is the 15° C. calorie. — S

* It is the fact that we are enabled to express $-d\theta/dt$ as a simple function of θ which gives the method its power. But it must not be assumed that we claim any generality for the five-fourths-power law. We say no more than that it holds with very considerable exactness for the liquids which we have tested over the range 20° to 50° C. It may be necessary to search for a function of another form in other tests. The value of K with our particular calorimeter varies, of course, with the nature of the liquid in the calorimeter.

§3. EXPERIMENTAL ARRANGEMENTS

The simple arrangement adopted is shown in figure 1, which is almost self-explanatory. The copper calorimeter *A*, which is supported as shown, is 14 cm. from the base to the flange, has an internal diameter of 10 cm. and is about 0·45 mm. thick. The width of the flange is about 1·25 cm. The lid *B*, to which is attached the coil *C* with its current leads *D* and potential leads *E* and the stirrer *F*, is screwed down to the flange during the course of an experiment, a thin film of grease being placed on the flange beforehand. *G* is a casing containing the ball bearings for the stirrer. The thermometer is inserted at *H*.

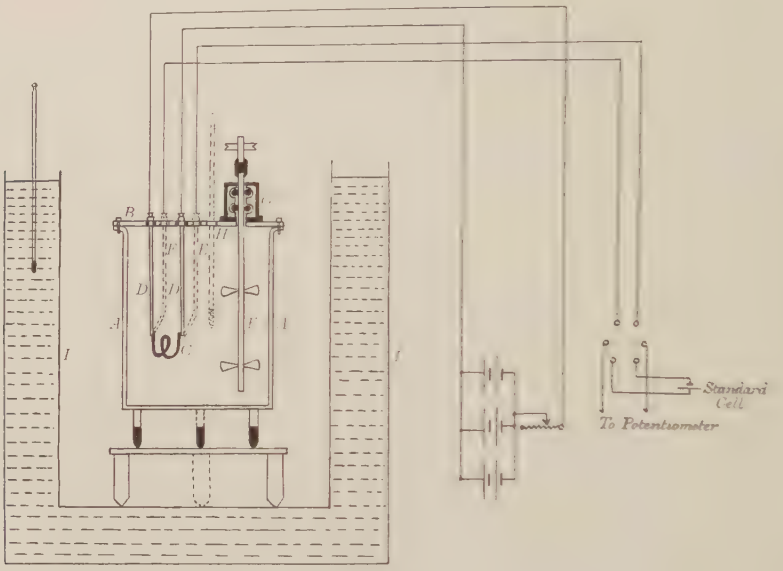


Figure 1.

The calorimeter is surrounded at a distance of about 8 cm. from its walls by a cylindrical double-walled tank *I* which contains water at room-temperature and serves to define the "temperature of the surroundings." In any one run this temperature never varied by more than 0·03° C. The stirrer was motor-driven at such a rate as to cause no appreciable rise of temperature during the period of an experiment. Nevertheless, the stirring was sufficiently vigorous to ensure appreciable uniformity of temperature throughout the mass of liquid. This was tested by systematically exploring the liquid by means of a sensitive thermocouple. Whatever the relative positions of the junctions, no difference of temperature so great as 0·1° C. could be detected at any stage of the heating.

At one stage of the proceedings, it was deemed that an exact knowledge of the temperature of the radiating surface of the calorimeter would be necessary, and experiments were made to test the temperature-gradient through the calorimeter

walls by soldering thermo-junctions inside and outside. It was found that a slight difference of temperature was established which increased with the temperature-excess of the contents of the calorimeter as shown in table 1.

Table 1

Temperature-excess (° C.)	2·90	9·27	12·86	18·43	22·87	29·61
Temperature-difference (° C.)	0·00016	0·00019	0·00021	0·00024	0·00029	0·00031

The thermometer used for reading the temperature of the liquid in the calorimeter was graduated in tenths from − 5° to 50° C. with an auxiliary scale at 100° C. It was an excellent instrument, supplied by B. Black and Son. It was standardized by comparison with a first-grade thermometer possessing a N.P.L. certificate.

The coil was of eureka wire and had a resistance of 0·1101 Ω. Its resistance was determined by comparison with a standard half-ohm coil. The temperature-variation of its resistance was neglected.

The potentiometer used was a simple Gambrell instrument and was calibrated in the ordinary way. Both the resistance-comparisons and the measurement of the potential difference at the terminals of the heating-coil were made on this potentiometer, which was standardized by means of a Weston cell.

Current was supplied to the heating-coil from a battery of secondary cells of capacity 50 ampère-hours, arranged in three rows in parallel, with two cells in series in each row.

§ 4. THE NATURE OF THE COOLING CURVE

It may be well to mention here that, although it is usually considered unwise to read temperatures on a falling thermometer, we have nevertheless found no difficulty in obtaining a very smooth experimental curve, and in reproducing our results to a high order of accuracy. This is probably to be attributed to the slight but continuous shaking produced by the stirring.

A typical curve is shown in figure 2 and is designated curve *A*. The ordinate scale for this is shown under *A* on the *y* axis.

As is well known, the rate of heat-loss ($-dQ/dt$) from a vertical wall, cooling under clearly defined conditions by natural convection, is proportional to the five-fourths ($\frac{5}{4}$) power of the temperature-excess. In this experiment, the calorimeter is losing heat very slightly by conduction, very slightly by radiation (this loss being minimized by highly polishing the external surface) and mainly by convection. We find, however, purely empirically, that the *rate of fall of temperature* ($-d\theta/dt$), and *not* the rate of heat-loss, is connected with the temperature-excess θ by the equation

$$-d\theta/dt = K\theta^{\frac{5}{4}} \qquad \ldots\ldots(3).$$

Whatever the régime, equation (1) holds at any stage of the cooling; and combining this with (3) we have, *under the conditions of cooling just specified*,

$$(MS + W)\, K\theta^{\frac{3}{4}} = E^2/RJ \qquad \ldots\ldots(4)$$

as the equation from which to determine the specific heat S at any temperature-excess θ.

We have now to consider how we establish the validity of (3). This leads by integration to

$$\theta^{-\frac{1}{4}} - \theta_0^{-\frac{1}{4}} = \tfrac{1}{4}Kt \qquad \ldots\ldots(5),$$

where θ_0 is the initial temperature-excess corresponding to the zero value of t. Experiment has furnished us with a series of values of θ and t. Hence if we plot

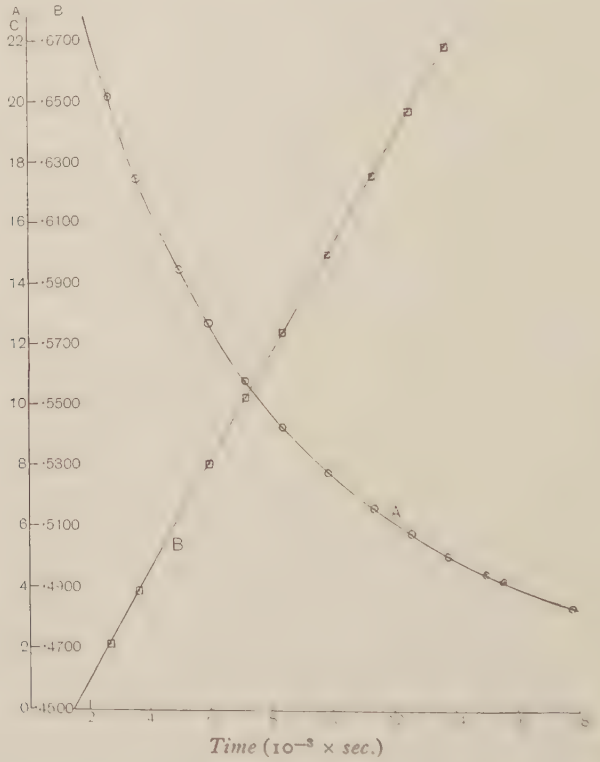


Figure 2. Ordinates of curve A are ° C., ordinates of curve B are $\theta^{-\frac{1}{4}}$.

$\theta^{-\frac{1}{4}}$ against t we obtain, if (3) holds, a straight line, the slope of which determines K. Such a line is the line B in figure 2; its ordinate scale is to be found under B on the y axis.

K may also be determined by a simple application of the rule of least squares to (5), and, in practice, we employ both methods. The graphical method, however, serves to test the equation as well as to determine K. It has been suggested that, in the very moderate temperature-range with which we are concerned, the form

of the law employed is a matter of secondary importance. It would not be difficult to devise a method by which, without assuming the $\frac{5}{4}$-power law, equation (3) might be tested by putting n for $\frac{5}{4}$ and determining its value experimentally. But we have thought it sufficient to assume the quasi-Newtonian form $-d\theta/dt = K\theta$ and to make a plot of $\log\theta$ against t, using our experimental data. The resulting line shows appreciable curvature, and indicates that the Newtonian law is certainly not applicable. The agreement between the observed and calculated values of θ using the $\frac{5}{4}$-law is shown for an experiment with aniline in table 2.

Table 2

t (sec.)	θ observed (° C.)	θ calculated from K_1 (° C.)	θ calculated from K_2 (° C.)
0	30·69	—	—
2700	20·15	20·11	20·14
3660	17·44	17·49	17·46
4500	15·48	15·53	15·51
6000	12·68	12·69	12·67
7200	10·76	10·84	10·82
8460	9·23	9·28	9·26
9900	7·77	7·81	7·80
11400	6·59	6·57	6·56
12600	5·75	5·76	5·75
13800	5·07	5·07	5·06
15000	4·49	4·48	4·47
15600	4·21	4·22	4·21
17800	3·36	3·38	3·37
18300	3·19	3·22	3·22
19500	2·88	2·89	2·89
22000	2·33	2·32	2·32

[K_1 found graphically $= \cdot 07004 \times 10^{-3}$; K_2 calculated by the method of least squares $= \cdot 07012 \times 10^{-3}$.]

§ 5. THE WATER-EQUIVALENT OF THE CALORIMETER

The water-equivalent of the calorimeter was determined by filling the calorimeter with water, and performing an experiment similar to that described. The experiment provides a fairly stringent test of the method, inasmuch as the water-equivalent is but about 12 per cent of the water-equivalent of the whole mass.

After carrying out the experiment as indicated, and using the equation

$$(MS_w + W)\, K\theta^{\frac{5}{4}} = E^2/JR,$$

we found that the water-equivalent increased linearly with the temperature. The *International-Critical-Tables* values of the specific heat of water S_w at the respective temperatures were used for calculating the results.

The relation between W and T can be expressed by

$$W_T = W_0\,(1 + \cdot 0005541),$$

where T is Centigrade temperature on the thermodynamic scale.

The agreement between the observed and calculated values of W is shown in table 3.

Table 3

Temperature T	22·05	26·33	33·31	38·32	40·65	45·25
W observed	71·22	71·39	71·67	71·87	71·99	72·21
W calculated	71·16	71·29	71·61	71·86	71·93	72·05

δ The calorimeter is almost completely of copper. δ, the specific heat of copper, is given very closely by the following relation*

$$\delta = \cdot 09088\,(1 + \cdot 0005341T - \cdot 00000048T^2)$$

between 0° C. and 100° C.

It will be seen that, although the water-equivalent of the calorimeter is but a small fraction of the whole water-equivalent, the method is sensitive enough not only to determine its value, but also to indicate fairly closely the magnitude of the temperature coefficient.

The water-equivalents used in subsequent calculations are, of course, those appropriate to the temperatures concerned.

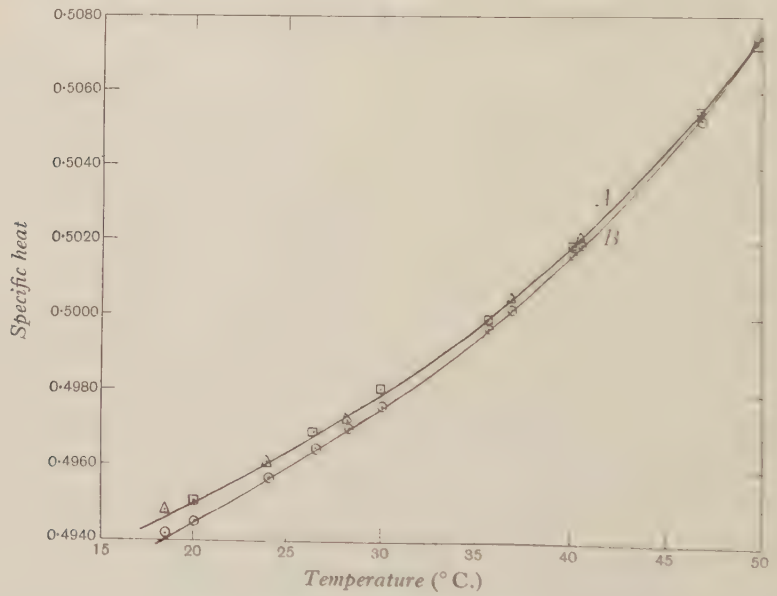


Figure 3.† Aniline: Ferguson and Miller's results, *A*; Lang's results, *B*.

* E. H. Griffiths and E. Griffiths, *Phil. Trans.* **213**, 119 (1914). *Proc. R. S.* **88**, 549 (1913); **90**, 557 (1914).

† Amended: see the discussion, p. 207.

§6. RESULTS FOR ANILINE

Although aniline has been widely used as a test liquid, it is by no means an ideal liquid on which to test a new method; it is hygroscopic and it polymerizes. The aniline used was of research quality and was dried over calcium chloride. It was then distilled and was collected in a closed flask which had previously been cleansed and dried by passing warm dry air through it for a considerable period. The calorimeter after being polished was similarly treated, and the aniline was transferred to the calorimeter by a siphoning process which involved practically no contact with undried air, the air which entered the storage flask as the aniline escaped being drawn over calcium chloride. The experiment was carried out as described, with the results as tabulated below in table 4. In this and other tables, T stands for Centigrade temperature, and θ for the temperature excess.

Table 4

(a) Heating experiment

Mass of liquid = 1099·30 gm.

E^2/R (W.)	Temperature of surroundings (° C.)	Temperature of calorimeter (° C.)	Temperature-excess (° C.)	Water-equivalent W	Specific heat S
0·6688	17·18	20·08	2·90	71·16	0·4951
2·8687	17·18	26·45	9·27	71·47	0·4969
4·3278	17·19	30·05	12·86	71·55	0·4981
6·8024	17·20	35·63	18·43	71·67	0·5000
8·9541	17·20	40·07	22·87	71·97	0·5019
12·4512	17·21	46·82	29·61	72·29	0·5054

[K_1 found graphically = $0{\cdot}07004 \times 10^{-3}$; K_2 found by the method of least squares = $0{\cdot}07012 \times 10^{-3}$.]

(b) Cooling experiment

Temperature of calorimeter (° C.)	Temperature of surroundings (° C.)	Time t (sec.)	Temperature of calorimeter (° C.)	Temperature of surroundings (° C.)	Time t (sec.)
47·90	17·21	0	22·95	17·20	12600
37·36	17·21	2700	22·27	17·20	13800
34·65	17·21	3660	21·69	17·20	15000
32·69	17·21	4500	21·41	17·20	15600
29·89	17·21	6000	20·54	17·19	17800
27·97	17·21	7200	20·38	17·19	18300
26·43	17·20	8460	20·07	17·19	19500
24·97	17·20	9900	19·52	17·19	22000
23·79	17·20	11400			

A second experiment gave almost identical figures for S, and these together with the values found by Dr Lang are shown in table 5 below.*

* In his paper Dr Lang's results are given in terms of the 20° C. calorie. In table 5, however, the values have been recalculated in terms of the 15° C. calorie. See the discussion, p. 207.

Table 5

Experiment 1

T	20·08	26·45	30·05	35·63	40·07	46·82
S (F. and M.)	0·4950	0·4969	0·4981	0·4999	0·5019	0·5054
S (H. R. L.)	0·4945	0·4965	0·4976	0·4997	0·5017	0·5052

Experiment 2

T	18·45	24·06	28·24	36·96	40·59	49·62
S (F. and M.)	0·4947	0·4961	0·4972	0·5005	0·5019	0·5073
S (H. R. L.)	0·4943	0·4957	0·4971	0·5003	0·5020	0·5074

Our results are very satisfactorily represented by the equation

$$S = 0{\cdot}495_0 + 0{\cdot}000275\,(T - 20) + 0{\cdot}0000035\,(T - 20)^2$$

and agree with those of Dr Lang to about 1 part in 1000. The results are shown graphically in curve A of figure 3, where the □'s and △'s represent our values for experiments 1 and 2 and the circles represent Dr Lang's values, the curve for these being designated B.

The values of the specific heat of aniline calculated from the formula just given are compared with the observed values in table 5*a*.

Table 5*a*

T	S observed	S from formula
20·08	0·4951	0·4950
26·45	0·4969	0·4969
30·05	0·4981	0·4982
35·63	0·5000	0·5002
40·07	0·5019	0·5021
46·82	0·5054	0·5053

§7. RESULTS FOR BENZENE

The benzene used was guaranteed free from thiophene or sulphur and was tested for freedom from thiophene by the fact that isatin dissolved in sulphuric acid showed no blue coloration when shaken with the benzene. Freedom from sulphur was indicated by the fact that no discoloration of mercury took place when the benzene had been shaken with concentrated sulphuric acid.

The benzene was fractionally crystallized, the portion employed in our experiments melting at 5·48° C. It was transferred to the calorimeter, with precautions similar to those just detailed, and the experiment was carried out in the manner

described. The results of two independent runs are given in table 6, and the values obtained by Williams and Daniels are appended for comparison.

Although not overtly stated, it may be inferred from Williams and Daniels' paper that the results are in terms of the 20° C. calorie. The results obtained by these two experimenters are shown in curve *B*, and our results on curve *A* of figure 4.

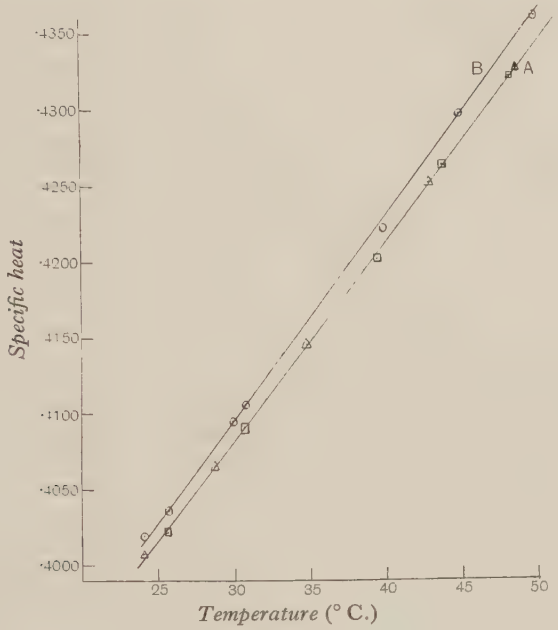


Figure 4. Ferguson and Miller's results, *A*; Williams and Daniels' results, *B*.

Table 6

Experiment 1

T	22·61	25·62	30·69	39·51	43·93	48·56
S (F. and M.)	0·3993	0·4025	0·4092	0·4220	0·4264	0·4321
S (W. and D.)	0·4015	0·4044	0·4106	0·4236	0·4280	0·4336

Experiment 2

T	24·03	28·65	34·80	42·83	48·89
S (F. and M.)	0·4008	0·4066	0·4147	0·4252	0·4325
S (W. and D.)	0·4022	0·4081	0·4162	0·4278	0·4340

The results, which on the average differ from those of Williams and Daniels by about 1 part in 500, are very fairly represented by the equation

$$S = 0{\cdot}395_8 + 0{\cdot}00125\,(T - 20).$$

The values of the specific heat of benzene calculated from the formula just given are compared with the observed values in table 6*a*.

Table 6*a*

T	22·61	25·62	30·69	39·51	43·93	48·56
S observed	0·3993	0·4025	0·4092	0·4220	0·4264	0·4321
S calculated	0·3995	0·4026	0·4093	0·4221	0·4265	0·4322

It would seem that the method is not without its advantages. It is easy in application, is fairly rapid, does not demand any excessive quantity of liquid, and may, with simple apparatus of the ordinary laboratory type, give results which are of accuracy greater than 1 per cent. Our own figures, we think, may be depended on to 1 part in 500 and we hope to improve the accuracy and range of our apparatus so as to obtain values for the specific heats of a number of important liquids over the range of their normal existence. The sensitivity of the method is considerably greater than we had hoped, and it has possibilities as one for determining, with precision, the variation with temperature of the thermal capacity per unit mass of water*. At least, it affords a laboratory method for the determination of Joule's equivalent which has some advantages over the standard electrical method of the text-books.

§8. ACKNOWLEDGMENTS

It is a pleasure to record our thanks to Dr Ezer Griffiths for his advice on certain technical matters, and to Prof. H. R. Robinson, who has afforded us every facility for the carrying out of the experiments.

DISCUSSION

Dr H. R. Lang. The authors have developed a simple calorimetric method that promises to give accurate results without difficulty. It is certainly a little disturbing to find that the usual elaborate precautions which are deemed necessary in calorimetric work of precision seem to have been omitted without any consequent appreciable decrease in accuracy. I am a little uncertain, however, about the efficiency of the stirring, and suggest that the authors should determine the temperature-gradient between the outside wall of the calorimeter and the centre of the liquid.

The comparison of results expressed in different units is surely not quite just. One does not think of comparing lengths in centimetres and inches; it is not so very different, I venture to suggest, to compare measurements in different calories.

* We have carried out some preliminary experiments to test this point, and have obtained values for the specific heat of water, over the range 20° to 50° C., in very close agreement with those obtained by Callendar and Barnes. We hope to extend the range of these observations in the near future.

I would therefore urge the authors to make the simple conversion, and to amend the tables and graphs before the paper is published in the *Proceedings*. It happens that the difference between the two calories accounts almost exactly for the divergence between the authors' results and those of other observers. The authors have omitted to include the units of specific heat in their tables. This, I assume, is an oversight.

Could the authors make a determination of the freezing-point of a sample of their aniline after an experiment? This would give a figure for the water content, and dispense with any assumption about its magnitude. I should be particularly grateful if the authors would make some measurements on aniline below 5° C. where my results were consistently peculiar.

Finally, may I suggest to the authors, that in view of some experiments that my colleagues and I have carried out during the past few years, when they make their experiments with water above 50° C. they should take the greatest care to ensure that it is quite air-free and that no air can dissolve in it during the experiment.

Dr Ezer Griffiths. This is an interesting development of the technique of the cooling-curve method of measuring specific heats of liquids. The authors have not referred to one useful feature of their method, namely that a single calibration will suffice for a series of different liquids, so that the method simplifies the problem of the heating-coil when corrosive fluids are dealt with.

As regards the calorie to be adopted the position is as follows:

A meeting of the International Union of Physics was held in Brussels in 1931. At this meeting discussion took place on matters brought forward by national committees. The British National Committee, represented by Sir Richard Glazebrook and myself, submitted a proposal as to the definition of the calorie. As a result of the discussion on this and other proposals a sub-committee was appointed to deal with symbols, units, and nomenclature in physics and to report to the General Assembly. At the first meeting of this sub-committee, Prof. Kennelly being in the chair, the following resolution was agreed to unanimously:

That this Commission recommends to the Executive Committee of the International Union of Physics that the following two propositions be communicated to the next General Assembly as recommendations from this Commission:

(1) That the unit of heat when measured in units of energy be the Joule, defined as equivalent to 10^7 ergs.

(2) That the gramme-calorie be the amount of heat required to raise the temperature of one gramme of water from 14·5° to 15·5° of the international scale of temperature.

Dr J. H. Brinkworth. I should like to ask whether the truth of the empirical law $d\theta/dt = K\theta^i$ was established for liquids having large temperature coefficients of specific heat. It seems that heat-losses of all kinds are effective in establishing this law, but the experimental conditions must be identical both during the

stationary-balance experiment and during the cooling. If during the latter the current used in the former is sent through one pair of potential and current leads, the conditions will be the same.

Mr J. H. Awbery. Of the various calorimetric methods which have been suggested, the one in which the temperature of a surface is used as a measure of the heat-flow from it, by means of a special calibration, has been perhaps the least employed. Prof. Lees used it about 36 years ago for conductivity-measurements, but he did not regard it as very accurate. To apply this method to specific-heat work is even more difficult, because in this case the flow is not steady. The authors therefore made a great advance when in one step they overcame this difficulty, and made the method a precise one. This they did by the device of fitting a curve of defined mathematical form to the cooling or heating curve. In this connexion, may I ask the authors whether, in addition to proving that $d\theta/dt = K\theta^{\frac{1}{2}}$, they also proved that $-dQ/dt = K'\theta^{\frac{1}{2}}$? It seems to me that, within experimental error, both might be true.

A second point to which I would refer is in table 1. It was valiant of the authors to try to measure such small temperature-differences, but surely the results only prove that the fall in temperature through the copper is negligible? Actually, each line of the table is a measure of the heat-flow and they hardly seem to bear out the $\theta^{\frac{1}{2}}$ power law. One quantity changes twofold when the other changes tenfold.

The suggestion of using the method to study the heat-capacity of water is a very attractive one, and I doubt not that I shall be like many others in looking forward to the results.

Mr J. H. Coste. The method described is interesting as yielding the specific heat at a defined temperature, rather than, as is usually the case, over a small range of temperature. It is not quite clear from the description of figure 1 how much of the adjuncts of the calorimeter vessel are included in the water equivalent, or of what material the lid, for instance, is made; but as the calculated and observed values agreed so well the question is one of interest rather than importance.

By their opening reference to organic liquids which play an important part in technological processes the authors may be considered to justify the statement that the method does not require any excessive quantity of liquid, but the volume used seems to have been rather more than a litre, which in many cases of only scientific interest would be a large amount to prepare. The value of the data for aniline would be more securely established if it could be stated that the dry liquid boiled within certain limits, or if other more precise criteria than the statement that it was of "research quality" could be given.

For an experiment where much depended on the temperature of the surroundings the means used to define this seem rather crude, as the thermometer shown in figure 1 only gives the temperature of the unstirred water in its vicinity, whereas the relevant temperature is that of the inner surface of the tank. Probably the mass of water in the jacket was very large and the room temperature very constant,

but a statement to this effect would be useful. The method appears to be capable of considerable precision and to be convenient. It would be of interest if the authors would use it to determine the specific heats of other pure organic liquids.

Authors' reply. We desire to thank those who have contributed to the discussion for their very helpful comments. To the order of accuracy contemplated, we found the temperature-gradient to be negligible between all the very widely distributed points selected. We have adjusted the results obtained by different experimenters so that in the corrected pages all results are expressed in terms of the same calorie; but in the matter of specific-heat units we are proof against Dr Lang's wiles, and refuse to be drawn into a discussion on that point. Dr Ezer Griffiths' remarks show very clearly the relations between the units involved. We shall bear in mind his suggestions in an extended series of observations which we may undertake.

In reply to Dr Brinkworth's remarks: The empirical law assumed has been tested for the liquids specified in the paper. It may happen that, with other liquids, a different form of $f(\theta)$ must be assumed, but this will in no way invalidate the principle of the method. As Dr Brinkworth points out, the conditions in the "statical" and "dynamical" experiments must be identical. In experiments with different liquids K varies, of course, from liquid to liquid, even if the five-fourths-power law holds.

No; we did not attempt to test whether $-dQ/dt = K'\theta^{\frac{5}{4}}$. We propose to make a study of the rate of heat-loss a separate problem. We thank Mr Awbery for his kindly description of our attempt to find the temperature-gradient through the cylinder as "valiant." Remembering a well-known line of Pope, we might well choose another epithet to describe our effort. As Mr Awbery quite correctly remarks, it simply shows that the temperature-drop through the copper is negligible.

We thought that the method of determining the water equivalent showed clearly enough how much of the calorimeter and its adjuncts entered into the water equivalent. The material of the calorimeter was practically all copper, save for a brass flange to which the lid was attached. The liquids used were bought as of "research quality" and were dried, and fractionally distilled or crystallized in the ordinary way. As Mr Coste correctly surmises, the mass of water in the enclosing tank was very large, and, as the experiment was carried out in a secluded corner of the laboratory, constant temperature or, rather, *very* slowly varying temperature conditions were closely fulfilled.

We hope to carry out experiments over a wide range of temperature, using a great variety of pure liquids.

551.594.11

THE IONOSPHERE AS A DOUBLY-REFRACTING MEDIUM

By Prof. E. V. APPLETON, F.R.S.

AND

GEOFFREY BUILDER, B.Sc.,
King's College, London

Received November 17, 1932. *Read December* 16, 1932.

ABSTRACT. In a previous communication to the Society the occurrence of wireless echo doublets was described and was provisionally attributed to the influence of the earth's magnetic field on the dispersive properties of the ionosphere. A more extended study of the subject, which has included an experimental determination of the polarizations of the doublet-components, has confirmed this hypothesis. In South-east England, for ionospheric reflection at vertical incidence, the echo-component of lesser delay is, in general, of right-handed, and the component of greater delay of left-handed, circular polarization, but this temporal sequence should be reversed in the Southern Hemisphere, and in certain special circumstances in the Northern Hemisphere.

§ 1. INTRODUCTORY

In the experimental study of the ionosphere* by electric wave reflection it is very frequently found that the reflected signal resulting from a single incident pulse consists of a doublet, the components of which are separated by a small time-interval. Appleton and Builder†, who described this phenomenon, suggested that it was possibly due to the influence of the earth's magnetic field, the two components being, in fact, the two elliptically polarized waves of opposite rotational sense into which the incident wave is split by magneto-ionic double refraction. Much further work has been since carried out on this subject, the result of which provides strong support for this interpretation. Such work forms the subject of the present communication.

§ 2. THE EXPERIMENTAL FACTS

The first curve given by Appleton and Builder, in which the equivalent height of the upper reflecting region was plotted as a function of time, showed that as the night advanced the temporal separation of the components of the echo gradually increased. Owing to electron-limitation, reflection eventually ceased to be marked, but it was found that the echo-component of greater delay was lost, owing to this cause, earlier during the night than was the echo of shorter delay. On the incidence of sunrise, however, the re-appearance of the echo of shorter delay was observed

* This is a convenient term, suggested by Mr R. A. Watson Watt, to connote the whole of the ionized regions of the upper atmosphere.

† *Proc. Phys. Soc.* **44**, 76 (1932); see also Rukop and Wolf, *Z. f. Tech. Phys.* **13**, 132 (1932); and White, *Nature*, 129, 579 (April 1932).

almost half an hour before the other appeared. This evidence suggested very strongly that a greater concentration of electrons is required to return the longer delay component than is required for the shorter delay component.

Various other experiments of the same kind have amply confirmed this interpretation, but perhaps another way of demonstrating the point is even more conclusive. One of the most useful indices of ionospheric conditions is a curve illustrating the relation between equivalent height of reflection and electric wave frequency. An example of such a curve is shown in figure 1. The data for it were obtained in the period 2353 G.m.t. on April 29 to 0020 G.m.t. on April 30, 1932, in experiments carried out between East London College and King's College, London. It will be seen that the phenomenon of splitting is more marked the higher the frequency, and that the F_1'' component penetrates the layer and is lost for a lower

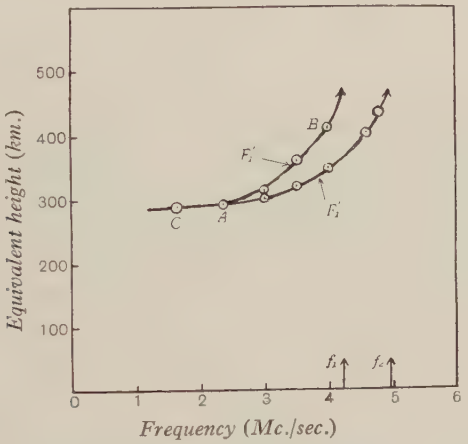


Figure 1. Curve illustrating relation between equivalent height and frequency. F_1' extraordinary; F_1'' ordinary ray.

critical frequency than is the case for F_1'. The curve also suggests that even for the lower range of frequencies the phenomenon of splitting actually occurs, but, because of the small temporal separation, the components are not resolved. This point could perhaps be settled using signal pulses of very short duration indeed. It may, however, be mentioned that for conditions represented by, say, *A* in figure 1 the essentially double nature of the echo signal is often indicated by marked intensity-variations in the central part of the pulse, the variations of the beginning and end being by no means so marked. These variations of intensity are naturally explained as being due to interference between the two parts of the components which overlap in time, and the phase difference of which is gradually altering.

Evidence derived from instances of multiple reflection also points very definitely to an interpretation in terms of the magneto-ionic theory. When the doubling of signals also occurs, the type of record obtained is that shown in figure 2 which is reproduced from our paper on wireless echoes of short delay. Here the received signal is shown as a function of time, the time scale being indicated by the lower

trace of frequency 1115 cycles per second. The ground signal G is followed by the component echoes F_1' and F_1'' of the first order and later by the corresponding second order reflections F_2' and F_2''. Now the fact that F_1' and F_1'' on their return to the upper atmosphere are not split again indicates that their polarization must be different from that of the waves incident on the layer at the first reflection. That they are reflected without further doubling does, in fact, appear to indicate that their polarizations are those which travel without change within the ionized medium. Also the fact that the temporal interval $F_1'F_2'$ is less than the corresponding interval $F_1''F_2''$ for the other component shows that F_1' and F_1'' must be differently polarized. Observations on the relative intensities of the two echo-components made from a large number of records have enabled us to make the following statements. For conditions such as are indicated by B in figure 1 the component of greater delay is of lower amplitude than the component of lesser delay. This is presumably due to the effects of electron-limitation in causing impaired reflection. On the other hand, as conditions represented in point A in

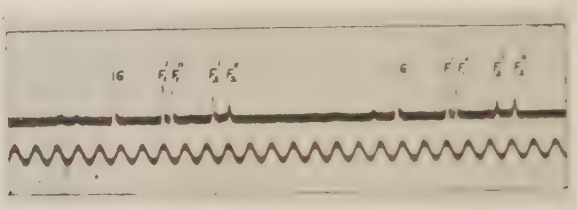


Figure 2. (The time scale shown under the record is that provided by an alternating current of 1115 ~.)

the same figure are approached, the intensity of the F_1'' component relative to F_1' increases and ultimately becomes the greater. This shows that the polarization for conditions represented by a point such as C is determined by the polarization of the stronger component F_1''. Now in some unpublished experiments carried out in 1930, using the frequency-change method of determining polarization, it was found that, for conditions in which splitting was not marked and for conditions such as may be represented by point C in figure 1, the polarization of 100-metre waves reflected from the upper region was approximately circular and left-handed*. The combined evidence therefore indicates that the polarization of F_1'' is left-handed and that of F_1' right-handed.

Although our early observations on echoes, made in the summer of 1931, showed the phenomenon of splitting in its simple form, the returned echo being in the form of a doublet, much more complicated results were obtained during the winter of 1931–2 to be followed again by simpler cases in the summer of 1932. An example record of one of the more complicated types of splitting is shown in figure 3. It represents a case of reflection from the upper region, and was obtained at 0226 G.m.t. on November 26, 1931, with a wave-length of 90 metres.

Although patterns of this type are sometimes so complicated as to be difficult

* The observer is here supposed to be looking in the direction of propagation.

to unravel, it is always clear that the first part of the returned signal fits exactly on the F_1' portion of the (P', f) curve*. This fact strongly suggests that the phenomenon is magneto-ionic in origin and that the first part of the complicated signal possesses normal characteristics and is therefore of right-handed polarization, the second and more complicated portion of the signal being associated with the left-handed component of greater delay. At the same time we may note that the more complicated type of splitting occurs only under conditions (e.g. winter and/or pre-dawn) when the ionization-gradient of the upper region is low. Further reference to this point is made below.

In attempting to settle definitely the actual polarization of the two supposed magneto-ionic components, experiments were carried out with a new type of polarization analyser†. It will at once be seen that the problem of determining the polarization of two signal impulses which follow one another with a very short time-interval is quite different from that involved in previous polarization investigations in which cases the frequency-change method and a sustained signal have been employed. For our present purpose it is not sufficiently convenient to

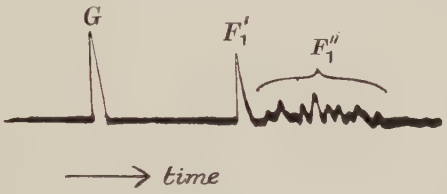


Figure 3.

determine the resultant polarization of either the combined effect of ground waves and downcoming waves or even of the resultant polarization of the two components on a suppressed ground-wave receiving system. The more straightforward way is to determine separately the polarizations of ground wave and of the echoes as they successively occur in time. This demands an indicator of very rapid response, such as the cathode-ray oscillograph.

In the new type of radio-polarimeter, which has been developed at Radio Research Station, Slough, two similar frame aerials are used at right angles and the radio-frequency signals received on them similarly amplified, without change of relative phase, and applied to the two pairs of deflecting plates of a cathode-ray oscillograph. In this way the cathode-ray trace gives the relative amplitudes and polarizations of the ground signal and the subsequent echoes in succession. Since the pulses are emitted by the sender at a regular rate (50 per second) it is possible, by means of a suitably synchronized time base, to make the traces corresponding to the ground wave and echoes quasi-stationary on the oscillograph screen so that they may be examined by eye.

* This is a curve in which the equivalent path is exhibited as a function of the electric wave frequency.

† The basic idea underlying this method of polarization delineation was independently suggested by Mr R. A. Watson Watt and one of the writers. A very brief account of it has been given in the issue of *The Wireless World*, July 8, 1932, p. 17.

It was found that when the polarimeter was in adjustment the trace corresponding to the ground-wave signal was a straight line, indicating that its magnetic vector was linearly polarized. When downcoming waves from the upper region were also received it was found that in the day-time, when splitting was not in evidence, the echoes were of variable elliptical polarization with a predominantly left-handed rotational sense. When the echo consisted of a normal doublet, it was found that the components were approximately circularly polarized, the component of lesser delay being right-handed and that of greater delay left-handed*. (It was possible at times to note this effect up to echo doublets of the fifth order.) These experiments therefore confirmed directly our first interpretation of the nature of split echoes.

§ 3. THEORETICAL EXPLANATION AND DISCUSSION OF RESULTS

(*a*) *The magneto-ionic theory of wireless propagation.* In a discussion on "Atmospheric ionization" before the Society in November 1924† it was pointed out by one of the authors that if the conductivity of the upper atmosphere is due to electrons (and not to ions of molecular mass) the propagation of wireless waves in the ionosphere is very considerably influenced by the earth's magnetic field. Gyroscopic terms in the dispersion equations, not considered in the ionic refraction theories of Eccles‡ and Larmor§, become of importance, and many of the phenomena familiar in magneto-optics are to be expected. This magneto-ionic theory, as the revised treatment was called, has been developed mathematically, for the general case of propagation at any angle relative to the earth's field, by Appleton ‖, Goldstein¶ and Hartree**.

On the experimental side Appleton and Ratcliffe†† have found that downcoming waves 400 metres in length are approximately circularly polarized with a left-handed sense of rotation in England, whereas Green‡‡ has shown that, under exactly comparable conditions in the southern hemisphere, the downcoming waves are of opposite rotational sense. Such results have been explained as being due to the differential absorption of the two components of opposite rotational sense which travel in an ionized medium under the influence of the earth's magnetic field, the absorption of the right-handed component being much greater than that of the left-handed component for waves of 400 m. travelling in a direction making an acute angle with the imposed field (e.g. for downward propagation in England). But according to the theory already developed, this differential absorption of the magneto-ionic components should be the less marked the shorter the wave-length,

* The rotational sense was found by advancing or retarding the phase of the signals received on one of the loops (by altering slightly the circuit tuning) and noting the effect on the circular traces on the oscillographic screen.

† E. V. Appleton, *Proc. Phys. Soc.* **37**, 22 (1925); see also Nichols and Schelleng, *Bell System Technical Journal*, **4**, 215 (1925).

‡ *Proc. R. S.* A, **87**, 79 (1912).

§ *Phil. Mag.* **48**, 1025 (1924).

‖ *Proc. U. R. S. I.* (Washington, 1927).

¶ *Proc. R. S.* A, **121**, 259 (1928).

** *Proc. Camb. Phil. Soc.* **27**, 143 (1931).

†† *Proc. R. S.* A, **117**, 576 (1928).

‡‡ *Nature*, **128**, 1037 (1931).

so that for short waves we might expect to observe them both with comparable amplitude.

The problem is perhaps best illustrated by taking a practical case of the reflection of, say, 80-metre waves at vertical incidence. We consider the propagation of the waves vertically downwards at an angle of $(\frac{1}{2}\pi - \delta)$ to the magnetic field, δ being the angle of dip. For such a case the relation between the imposed field, its components, and the direction of propagation are as illustrated in figure 4. Here the magnetic field components, transverse and longitudinal to the direction of propagation, are represented by H_T and H_L respectively. H is the total magnetic force. δ

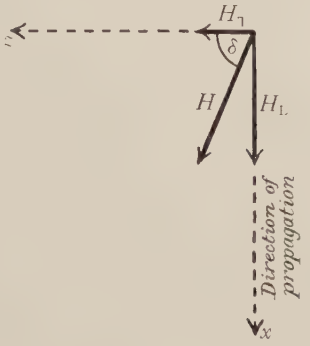


Figure 4. (The y axis is at right angles to the paper.)

For such conditions the dispersion equation relating the refractive index μ to the angular frequency p may be written μ p

$$\mu^2 = 1 - \frac{2p_0^2}{2\left(p^2 + \frac{1}{3}p_0^2\right) - \frac{p^2p_T^2}{p^2 - \frac{2}{3}p_0^2} \mp \sqrt{\left(\frac{p^4p_T^4}{(p^2 - \frac{2}{3}p_0^2)^2} + 4p^2p_L^2\right)}} \quad \ldots\ldots(1),$$

where $p_0^2 = 4\pi Ne^2/m$; $p_T = H_Te/mc$; $p_L = H_Le/mc$. Here N is the number of electrons, of charge e and mass m, per cm.^{3} and c is the velocity of electromagnetic radiation in free space. This equation is valid for negligible or very small friction. The formula with the upper sign refers to the extraordinary ray and that for the lower sign to the ordinary ray. It should be noted that p_0^2 is proportional to the ionization-content N. The corresponding formula for the polarization is N e, m, c

$$\frac{H_y}{H_z} = \frac{ipp_L(1 - \mu^2)}{(1 - \mu^2)\left(p^2 + \frac{1}{3}p_0^2\right) - p_0^2} \quad \ldots\ldots(2),$$

where the appropriate value of μ^2 is inserted from (1).

A curve illustrating the relation between μ^2 and N for a case of the propagation of 80-metre waves vertically downward in South-east England has been kindly drawn for us by Miss A. C. Stickland of the Radio Research Station, Slough, and is reproduced in figure 5.

(*b*) *The two types of magnetic splitting and their occurrence.* Now in the pulse measurements in which the doubling of echo signals has been most studied, we are concerned with the separate arrival in time of the two components. Such a group separation of the components can readily be seen to occur in two ways.

We see from the above curve relating μ^2 to N that the refractive index of the extraordinary ray is diminished to zero for a lower value of N than is the case with the ordinary ray*. This means that, for a layer the ionization of which increases with height, the reflection of waves at vertical incidence will take place at a lower level for the extraordinary ray than for the ordinary ray. Because of this we should expect echo pulses associated with the ordinary ray to arrive back at the ground a little behind the other component in time. Such a result is found in the most common type of splitting, and since it arises from the way in which the ionization in the upper atmosphere is stratified we call it "stratification splitting."

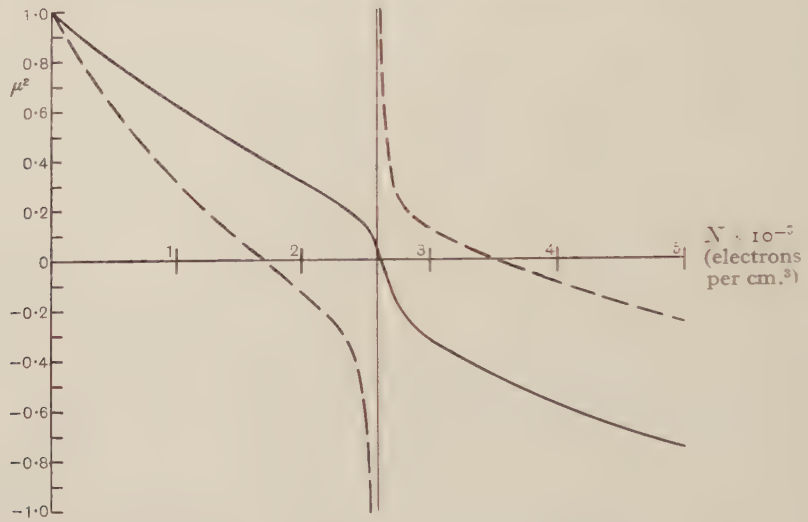


Figure 5.

——— extraordinary ray (μ_2)

——— ordinary ray (μ_1)

(N, μ^2) curves for $\lambda = 80$ metres. (Infinity of μ_2 at $N = 2{\cdot}58 \times 10^5$. Zero of μ_1 at $N = 2{\cdot}62 \times 10^5$.)

There is, however, another possible way in which the separation of the two components may take place. It is known that there are two ionized regions in the upper atmosphere and that, owing to the fact that the ionization in the upper region may be more intense than that in the lower region, waves which have penetrated the lower region may still be reflected by the upper region. In passing through the lower region, however, any signal pulse is retarded owing to the fact that the group velocity in an ionized medium is less than that in free space. In the case of the two magneto-ionic components, moreover, such group retardation will be differential, and it may be shown from the dispersion formula given above that the group velocity of the extraordinary ray is, for the special conditions in which we are interested, less than that of the ordinary ray. The effect of electron ionization in the Kennelly-Heaviside layer (lower region) is thus to make the extraordinary

* For the extraordinary ray μ^2 first becomes zero when $p_0^2 = \frac{3}{2}(p^2 - pp_H)$, while the same condition is reached for the ordinary ray when $p_0^2 = \frac{3}{2}p^2$.

component lag in time behind the ordinary component. "Group-retardation splitting," as this type of separation may be termed, therefore tends to produce the opposite effect to that produced by "stratification splitting," for stratification splitting retards the ordinary component behind the extraordinary component while group-retardation splitting produces just the reverse effect.

It has already been noticed that the group-retardation effect due to the ionization in region E suffered by waves which are ultimately reflected by the upper region F is most marked for frequencies just above the critical range of values for which penetration of the lower region begins to take place. We should therefore expect group-retardation splitting to occur most readily for these same experimental conditions. This has actually been found to be the case in experiments carried out with the radio-polarimeter, and these conditions are, as a matter of fact, the only ones under which we have observed splitting of this type. More usually the effect of group-retardation splitting is more than neutralized by that of stratification splitting, so that the first echo-component experienced is right-handed and the second component left-handed in rotational sense.

The dispersion curves illustrating the variation of μ^2 with N immediately show that a greater value of electron-concentration is required for the ordinary ray than for the extraordinary ray to produce the condition $\mu^2 = 0$ which we associate with reflection at vertical incidence. We have, in early morning tests, repeatedly found examples of this. The ionization is at that time increasing owing to the influence of the sun, and we find that, after electron-limitation has been causing the absence of echoes during the night, the shorter delay component is first noted and is then followed a few minutes later by the appearance of the second component. After this had been observed we were able to change to a shorter wave-length and see the process repeated.

In our attempts to determine the frequencies which just penetrate the upper region at different times of the day, we have often had occasion to obtain data from which a curve illustrating the relation between the equivalent path of the atmospheric waves P' and the frequency f could be drawn (see for example figure 1). These curves show that for a given value of N the ordinary ray of one frequency, say f_1, just penetrates the upper region while, at the same time, the extraordinary ray of another frequency f_2 has reached similar critical conditions. Now while, as has previously been mentioned, it is difficult to say exactly what are the deciding factors which determine reflection, we are fairly clear that the conditions for reflection must approach those indicated by $\mu \to 0$. But for frequencies greater than the characteristic gyro-frequency f_H (i.e. $He/2\pi mc$) the condition $\mu \to 0$ is reached for the extraordinary ray of frequency f_1 when

f f_1 f_2 f_H

$$N = \frac{3}{2}\frac{\pi m}{e^2}(f_1^2 - f_1 f_H) \qquad \ldots\ldots(3),$$

whereas the same condition is reached for the ordinary ray of frequency f_2 when

$$N = \frac{3}{2}\frac{\pi m}{e^2}(f_2^2) \qquad \ldots\ldots(4).$$

The two frequencies must thus be connected by the relation

$$f_2^2 = f_1^2 - f_1 f_H$$

or

$$(f_1 - f_2) = f_H\left(\frac{f_1}{f_1 + f_2}\right) \qquad \text{.....(5).}$$

Since the frequency-difference $(f_1 - f_2)$ is small we see that it is approximately equal to $\frac{1}{2}f_H$ or to about 0·70 Mc. sec. Now it is one of the striking facts of the experimental results that in all the (P', f) curves we have drawn in which the electron-limitation effect for the two components is shown, the frequency-difference corresponding to $(f_1 - f_2)$ is always of this order of magnitude, the mean experimental value for a large number of curves being 0·80 Mc. sec. We believe therefore that this agreement between theory and experiment is additional evidence in favour of the magneto-ionic theory.

The application of the magneto-ionic theory to the explanation of phenomena associated with the reflection of waves from region E has not proved so easy of interpretation as in the case of region F. It is known that, in general, waves reflected from the lower region are predominantly of left-handed polarization, and this has been attributed to the more marked absorption of the right-handed component during the passage of the waves through the lower and more attenuating part of the ionized region, which is situated below that at which the waves are reflected. The question whether the refractive process causes separation of the ordinary and extraordinary rays by stratification splitting in region E, in a manner similar to that found in the case of region F, is more difficult to settle. The ionization-gradient in region E appears to be greater than that in region F, so that simple stratification splitting as indicated by the temporal separation of the two components does not show up so readily. It can, however, be stated that splitting which appears, in the ordinary cathode-ray oscillographic delineation of echoes, to be of exactly the same type as that experienced for region F has been noted on many occasions, but we have not yet been able to make definite determinations of the sense of polarization of the two (or more) components separately. Certain phenomena noted under special circumstances, however, assume importance in this connexion. As has been described for reflection from region F, the effect of electron-limitation in a case in which the ionization is steadily decreasing is to cause the polarization to change from the predominantly left-handed rotational sense (when the polarization is determined by the greater amplitude of the ordinary ray) to the right-handed sense (when, owing to electron-limitation, the ordinary ray has penetrated the layer and only the extraordinary ray is being reflected). During sunset runs we have also sometimes found this to be the case for region E. Using a constant frequency the polarization of waves has at first been left-handed and later, as the ionization has decreased, has become right-handed. As the ionization has further decreased even the extraordinary ray has finally no longer been reflected, and full penetration of the lower region has taken place. Because of these magneto-ionic effects in cases of reflection from the lower region, and of the existence of group-retardation splitting for waves reflected by the upper region, we conclude

that the Kennelly-Heaviside layer contains a considerable proportion of free electrons.

(*c*) *The effects of friction.* The expression (1) above for the refractive index of an ionized medium under the influence of an imposed magnetic field was derived for assumed non-dissipative conditions. When account is taken of the collisional friction suffered by the vibrating electrons the value of the refractive index cq, which is now complex, is given* by

cq

$$(cq)^2 = \left(\mu - \frac{i\kappa c}{p}\right)^2$$

$$= 1 - \frac{2p_0^2}{2\left(p^2 + \frac{1}{3}p_0^2 - ip\nu\right) - \frac{p^2 p_T^2}{p^2 - \frac{2}{3}p_0^2 - ip\nu} \mp \sqrt{\left(\frac{p^4 p_T^4}{(p^2 - \frac{2}{3}p_0^2 - ip\nu)^2} + 4p^2 p_L^2\right)}} \quad \ldots\ldots(6).$$

where μ is now the real part of the refractive index, κ the absorption coefficient and ν the frequency of the electron collisions with gas molecules. As in (1), the upper sign refers to the extraordinary ray and the lower sign to the ordinary ray.

μ, κ

ν

The corresponding formula for the polarization is

$$\frac{H_y}{H_z} = i\left\{\frac{2p_L\left(p^2 - \frac{2}{3}p_0^2 - ip\nu\right)}{pp_T^2 \pm \sqrt{\{p^2 p_T^4 + 4p_L^2\left(p^2 - \frac{2}{3}p_0^2 - ip\nu\right)^2\}}}\right\} \quad \ldots\ldots(7).$$

The two formulae (6) and (7) are therefore the most general representative equations of the magneto-ionic theory in terms of a ray treatment, and their elucidation and application constitute one of the chief problems of present-day wireless propagation theory. We can here only mention a few of the more salient deductions which can be made from them.

In the first place (6) and (7) yield, as they should, formulae identical with those given by Lorentz for the special cases of transmission longitudinal and transverse to the imposed magnetic field. They also show that, for vertically downward propagation in South-east England at an acute angle with the earth's magnetic field, as in our present problem, the left-handed component is less absorbed than the right-handed component.

When friction is taken into account it is also possible to explain a difficulty which arises in the elucidation of the formulae obtained when friction is neglected. For the purely longitudinal type of propagation it is found from (1) that the refraction index of the extraordinary ray becomes zero when $p_0^2 = \frac{3}{2}(p^2 - pp_H)$, while the same condition is reached for the ordinary ray when $p_0^2 = \frac{3}{2}(p^2 + pp_H)$. Both ordinary and extraordinary rays are, moreover, in this case circularly polarized for all values of p_0^2. For all other directions of propagation, however, as has been mentioned above, the refractive index of the medium becomes zero for the extraordinary ray when $p_0^2 = \frac{3}{2}(p^2 - pp_H)$ and when $p_0^2 = \frac{3}{2}(p^2 + pp_H)$, while the same condition is reached for the ordinary ray when $p_0^2 = \frac{3}{2}p^2$. In the propagation transverse to the magnetic field the linear polarizations are maintained for all values

* Appleton, *J. Inst. E. E.* 71, 642 (1932).

of p_0^2, but for intermediate directions the polarizations change as the waves travel into regions of different p_0^2. For very small and very large ionizations (that is for small and large values of p_0^2) the polarizations of both components are circular, but in the region from $p_0^2 = \frac{3}{2}(p^2 - pp_H)$ to $p_0^2 = \frac{3}{2}(p^2 - pp_H)$ rapid variations take place, the polarizations of both ordinary and extraordinary rays being linear when $p_0^2 = \frac{3}{2}p^2$. We therefore see that, when friction is neglected, longitudinal propagation stands out as a special case, in that there are only two values of p_0^2 for which the refractive index becomes zero whereas for all other directions there are three.

A more detailed examination of the general formulae (6) and (7), in which friction is taken into account, shows that there is no real discontinuity between the longitudinal and the other cases. It is found that the nature of the type of friction is determined by the value of $2p_L \nu p_T^2$, a quantity which depends on the direction of propagation, the collisional friction and the earth's magnetic intensity. When the value of this quantity is small compared with unity, the propagation resembles more closely the transverse type. The polarizations then become plane for the conditions $p^2 = \frac{2}{3}p_0^2$, while the refractive index for the ordinary ray approaches zero at the same point. On the other hand, when $2p_L \nu p_T^2$ is large compared with unity the propagation resembles more nearly the purely longitudinal type. The polarization does not then become plane for the conditions $p^2 = \frac{2}{3}p_0^2$, nor does the refractive index of the ordinary ray approach zero for the same conditions.

These considerations have a bearing on a question which has been discussed above. If the propagation, for conditions under which reflection takes place, resembles the transverse type, we have seen in equation (5) that the frequency difference $(f_1 - f_2)$ between the two frequencies which penetrate the region as extraordinary and as ordinary waves is equal to $f_H f_1 (f_1 - f_2)$. If, however, the critical ratio $2p_L \nu_l p_T^2$ is much greater than unity, it is easy to show that under these conditions, which approximate to longitudinal propagation, the frequency difference $(f_1 - f_2)$ now becomes equal to f_H or to 1·32 Mc. sec. The corresponding theoretical value for propagation resembling the transverse type, as has been shown, is about 0·70 Mc. sec., while the value obtained experimentally for the upper region is 0·80 Mc. sec. Our knowledge of the values of ν at various heights is somewhat uncertain, but there is little doubt that for the upper region, where ν must be numerically 10^3 or less, we are dealing with a case approximating to the transverse type. The experiments support this, since the experimental value of $(f_1 - f_2)$ is much closer to the transverse value of 0·70 Mc. sec. than to the longitudinal value of 1·32 Mc. sec. For the lower ionized region, where ν may be considerably higher, it does not appear possible yet to make any such corresponding decision with great certainty.

(*d*) *The limiting polarization of wireless waves as observed at the ground.* It seems reasonable to assume that the polarization of waves received at the ground is the same as that which they possess when travelling through the lowest part of the ionized region from which they have emerged. For comparison with our experi-

mental results, therefore, we require to find the polarization for small ionization. Putting p_0^2 equal to zero in (7) we find

$$\frac{H_y}{H_z} = i\left\{\frac{2p_L(p-i\nu)}{p_T^2 \pm \sqrt{\{p_T^4 + 4p_L^2(p-i\nu)^2\}}}\right\} \qquad \ldots\ldots(8).$$

Further simplification of this formula depends on whether we have $p \gg \nu$ or $p \ll \nu$ for the region in which ionization falls to a low value. In the former case we have

$$\frac{H_y}{H_z} = i\left\{\frac{2pp_L}{p_T^2 \pm \sqrt{(p_T^4 + 4p^2p_L^2)}}\right\} \qquad \ldots\ldots(9),$$

while in the latter case

$$\frac{H_y}{H_z} = \left\{\frac{2p_L\nu}{p_T^2 \pm \sqrt{(p_T^4 - 4p_L^2\nu^2)}}\right\} \qquad \ldots\ldots(10).$$

For our present case of the propagation of 80-metre waves vertically downwards in South-east England we have either in case (9) or in case (10) polarizations which are very nearly circular. It is not until propagation is nearly at right angles to the earth's magnetic field that we find the limiting polarizations tending to change over to the linear state. We thus find that the experimentally observed circular polarizations are such as are to be expected from the theory.

If we compare the limiting polarizations with the characteristic polarizations which obtain at the regions of higher ionization where the waves are reflected, we are able to explain how it is that the extraordinary ray component F_1' is usually less distorted than is the ordinary ray component F_1''. For all values of the quantity $2p_L\nu/p_T^2$, such as we have to deal with in vertical propagation in South-east England, the polarization of the extraordinary ray alters but little in its passage into and out of the ionized region. On entry it is circularly polarized, while, for the condition $p_0^2 = \frac{3}{2}(p^2 - pp_H)$ when it is reflected, we have $H_y/H_z = ip_L/p_H$, which means that the polarization is still not very far from circular. The extraordinary ray travels therefore almost unaltered in polarization, as in the purely longitudinal type of propagation. On the other hand, matters are quite different for the ordinary ray. When this ray enters the ionized region it is circularly polarized and, unless $2p_L\nu/p_T^2$ is large, its polarization alters rapidly with increasing ionization*. Such alteration can only take place as a process of re-splitting. Under conditions of low ionization-gradient such re-splitting will be attended with an increase in the temporal duration of the signal pulse. This result will also be emphasized by a process of partial internal reflection, which is easily seen to be possible between the regions at which the refractive index becomes zero. An additional, though perhaps not predominant, effect may be due to the fact that, when dP'/df is large, the component frequencies of the pulse itself are very much separated, with consequent distortion. When both components are present this effect will, in general, be greater for the ordinary than for the extraordinary ray. Departures from a smooth ionization-gradient doubtless add to the complexities of the pulses, while temporal changes in N and H cause the rapid intensity-variations of the distorted pulses which are experienced at almost all hours of the night.

F_1'
F_1''

* If $2p_L\nu/p_T^2 < 1$ the polarization changes from circular to linear.

§ 4. ACKNOWLEDGMENTS

The work described in this paper was done as part of the programme of the Radio Research Board of the Department of Scientific and Industrial Research. The experiments were carried out partly at King's College, London, and partly at the Radio Research Station, Slough, and we wish to express our indebtedness to Mr R. A. Watson Watt, the Superintendent of that station, for his kind collaboration in part of the experiments and also for his helpful advice and discussion at all stages of the work. We also wish to thank Mr J. F. Herd, Mr L. H. Bainbridge-Bell and Mr R. Naismith, of the Radio Research Station, for their assistance in the course of the experiments.

While the work has been in progress we have enjoyed the great advantage of being able to compare our results with those obtained by Messrs J. A. Ratcliffe and E. L. C. White of the Cavendish Laboratory, who were at the same time investigating similar problems by a different method which they had independently devised.

DISCUSSION

Mr R. A. Watson Watt. The authors have carried one important stage further the description and explanation of the increasingly complex phenomena of ionospheric reflection which are revealed by increasingly powerful observational methods applied over extending ranges of time-of-day and season. The clarification of the conditions for "stratification splitting" and for "group-retardation splitting" and the discussion of the relative complexity of the echo pattern for the ordinary ray-component F_1'' are specially valuable features of the present communication.

Mr R. H. Barfield. I think it would be of great interest to have some statement of the relative frequency of occurrence of the two types of echo pattern illustrated in figures 2 and 3 respectively. Is the possibility that scattering, as opposed to magneto-ionic splitting, might be an important cause of the phenomena of figure 3 completely excluded?

Authors' reply. In reply to Mr Barfield's first question we can only say that distorted echoes are almost always found under conditions where electron-limitation of amplitude is operative. This applies to both ordinary and extraordinary rays. As we have mentioned in the text, a low ionization-gradient favours the formation of complex multiplets.

The fact that all parts of a distorted echo show definite circular polarizations, of one sign or the other, appears to rule out any explanation in terms of scattering alone, the magnetic field of the earth being neglected. The complex echoes with their ceaseless amplitude-changes may be taken as expressing the disorderliness of the electron distribution. We do, however, believe that each part of a complex echo represents partial reflection from a part of the medium which is large compared with a vacuum wave-length, so that, to that extent, we rule out scattering as usually understood in optics.

621.396.2.029.1:551.594.11

THE GENERATION AND RECEPTION OF WIRELESS SIGNALS OF SHORT DURATION

By J. F. HERD, A.M.I.E.E., Radio Research Station, Slough

Communicated by R. A. Watson Watt, September 15, 1932. *Read December* 2, 1932.

ABSTRACT. In the use of short-duration signals for investigation of the ionosphere, it is convenient to emit the signal pulses at a controlled rate and to receive them on an oscillograph giving temporal discrimination between the arrival of the ground ray and that of subsequent echoes. Simplification of recording technique can be obtained if, at the receiver, an oscillograph of the cathode-ray type is used in conjunction with a time-scale voltage whose rate of recurrence is exactly the same as that of the emitted impulses. The paper describes methods of utilizing the common frequency of an a.c. supply network to secure such synchronization.

The arrangement is first described in connection with ticking-grid circuits at the emitter and as a linear time-base device at the receiver. Alternative linear time bases using mercury-vapour triodes (thyratrons) are then discussed, and these circuits are combined with known means of obtaining a circular time scale to produce a spiral time scale affording wide temporal separation of observed phenomena and more fully utilizing the whole oscillograph screen. Lastly an improved method of impulsing the transmitter is described, involving the use of a thyratron as an abrupt switching device.

§ 1. INTRODUCTION

THE use of wireless signals of short duration is already well known as a means of investigating the conditions of the ionosphere.

In a recent communication Prof. E. V. Appleton and Mr G. Builder* have described a simple method of generating radio-frequency impulses of short duration. The method utilizes the fact that in an oscillating triode generator, if a suitable condenser-resistance combination be inserted in the grid circuit as shown at CR in figure 1, the triode generates oscillations in the form of brief impulses alternating with periods of quiescence whose duration is of the order of 25 or more times that of the duration of the oscillatory train. The same circuit has also been used for the production of a time base for use with a cathode-ray oscillograph†. During the period of oscillation a voltage is built up across the condenser C abruptly driving the fluorescent spot to the left of the screen, when the spot proceeds to return from left to right in a slower leakage-time of the relative order stated. With a condenser-resistance combination the left-to-right travel occurs exponentially, but can be made linear by the use of a saturated diode in place of the ohmic resistance.

The reception of such short-duration impulses provides a method for determination of the time between the arrival of a ground-ray signal and that of one or more

* *Proc. Phys. Soc.* **44**, 76 (1932).

† *Proc. R. S.* A, **111**, 672 (1926).

signals returning from the ionized layer. In this application, it is convenient, as is stated in the paper by Appleton and Builder, to make visual observations on a cathode-ray oscillograph, with photographic recording with an oscillograph either of this type or of the electro-mechanical or Duddell type. For visual observation with the cathode-ray oscillograph it is desirable to hold the time base so that its recurrence-frequency agrees exactly with the pulse-frequency of the radiated oscillation. In this case the ground-ray signal always falls on the same place in the left-to-right travel of the fluorescent spot, with the indirect-ray signals following after it in each sweep according to their continuously varying conditions which are the subject of investigation. For photographic recording a convenient and economical method is to use the same general arrangement and take snap exposures at suitable intervals of the varying conditions of the indirect-ray signals.

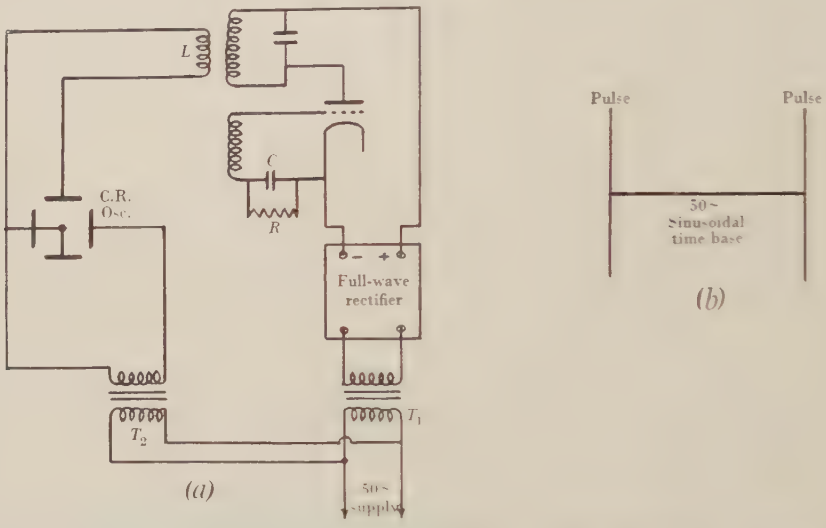


Figure 1. *a*. Method of locking transmitter to half-waves of voltage supply. *b*. Typical screen image when pulse-recurrence is locked to rectified half-waves of supply voltage.

With a suitable fluorescent material (e.g. calcium tungstate) on the oscillograph screen and by means of a camera of simple type, it is possible to obtain snap photographs with exposures of the order of from 0·2 to 0·1 sec., which give a record whose depth and quality are adequate for analysis. Moreover the use of the cathode-ray oscillograph for this purpose has the advantage that the same device serves both for monitoring and for photography. In addition, the monitoring indication remains visible during the recording process, while the photographic recording can be brought into instant operation either at regular intervals or at any special moment justified by the conditions.

For satisfactory photography, it is necessary that the image on the fluorescent screen should remain absolutely stationary. It is therefore essential that the recurrence-frequency of the time-base stroke should agree most accurately with the pulse-frequency of the transmitter. While it is possible to achieve a fair approach

to a stable pattern by manual control, e.g. at the receiver, some form of locking or master-control appears to be desirable. The use of frequency-controlled 50 ~ supply mains at both transmitter and receiver suggested itself as a means of impressing a stable frequency on the recurrence at both ends, and this paper describes simple methods which appear satisfactory for the purpose and have several points of novelty at both transmitter and receiver.

§ 2. CONTROL OF PULSE-GENERATION

The method consists essentially in using an unsmoothed supply, either from unrectified alternating current or from the unsmoothed output of a full-wave rectifier fed from the a.c. supply mains to energize the anode circuit of the valve and cause the signal pulse to recur at the same point in successive cycles of the supply voltage. The arrangement was first tried on a laboratory scale, in the conditions described below, and proved completely satisfactory in giving the desired locking state. Modifications for operating at higher power-levels are discussed later in the paper.

In order to utilize the higher frequency, namely 100 ~, both of signal pulses and of time-base recurrence, the second-mentioned arrangement (using unsmoothed rectified half-waves) was first tried. The local transmitter consisted of a simple coupled-circuit oscillator with grid condenser C and resistance R for control of the pulse-frequency. The usual smooth d.c. supply to the anode circuit was replaced by the unsmoothed supply from a full-wave rectifier, as shown in figure 1 *a*.

A convenient method of adjusting the transmitter is also shown in figure 1. Values of C and R were first set so as to give approximately the desired frequency-recurrence of 100 per second. A coil L loosely coupled to the oscillating circuit impressed the transient trains of oscillations on one pair of plates of a cathode-ray oscillograph. The other pair of plates was supplied with a timing voltage of 50 ~ by a small transformer T_2. It was then found quite easy to adjust the CR value (the variable element was actually the condenser C) until a perfectly steady picture of the type shown in figure 1 *b* was obtained. As synchronization was approached the picture fell into a pattern in which the impulses moved jerkily towards the outer ends of the time-base line showing increasing amplitude as the impulses approached the ends. On final adjustment the impulses fell exactly at the ends of the base-line, as shown in figure 1 *b* and were then of the maximum amplitude observed, showing, presumably, that the incidence of the oscillation-train agreed exactly with the peak of the supply voltage from the full-wave rectifier. With this adjustment, the picture remained locked for very long periods and there appeared to be no difficulty either in keeping it perfectly stationary or in reproducing the conditions from day to day.

§ 3. LOCKING THE RECEIVER TIME BASE

In the case of the time base at the receiver, the same general method was employed with complete success, viz. to feed the anode of the oscillating circuit by means of the unsmoothed output of a full-wave rectifier. The modifications made in the time-base circuits were very minor and consisted merely in substituting this

unsmoothed supply for the normal battery supply of the oscillator (after removal of a by-pass condenser that was actually installed), while still leaving the battery connected to the diode for the purpose of maintaining the improved linearity which this confers.

The behaviour of the time base was then somewhat peculiar, and a little explanation of the phenomena observed is called for. The condenser was set to a value which was known to give, in conjunction with a suitable diode brightness, a good measure of linearity in the main stroke. As a timing reference, the 50 ~ supply was fed to the vertical plates via a small transformer in order to study the behaviour of the horizontal time-base stroke. The diode was then started at a dull glow and gradually

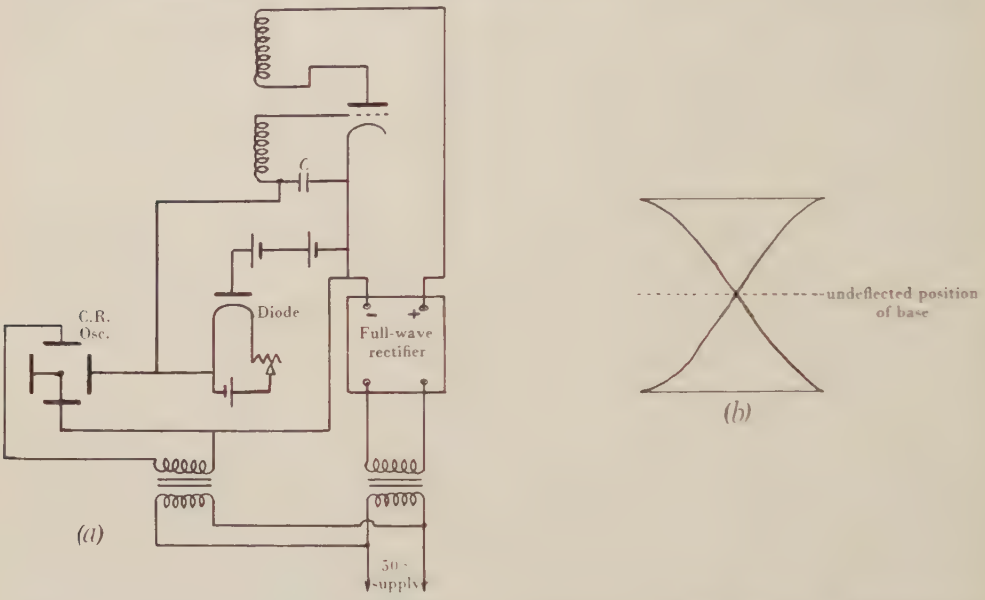


Figure 2. *a*. Method of locking receiver time base to half-waves of voltage supply. *b*. Typical screen image when time base is locked to rectified half-waves of supply voltage.

brightened. It was very noticeable that the pattern showed a marked tendency to lock on the sub-multiples of 50 ~ through which the time base then passed in the course of this adjustment. Thus it was very easy to obtain a stationary picture of 2 : 1, 3 : 2 and finally 1 : 1 when the time base was exactly of the duration $\frac{1}{50}$ sec. As the diode was continuously brightened, the picture tended finally to fall rather sharply into the pattern of 1 : 2 as shown in figure 2 *b*, which shows the time base of $\frac{1}{100}$ sec., breaking the 50 ~ voltage into its two half-cycles.

The setting which gave this stationary image of half-cycles was found to be the optimum for operation. Further brightening of the diode beyond this point introduced complicated effects and patterns which were extremely difficult to interpret and might be misleading. Fortunately there was no need ever to use these, since the arrival at the condition indicated was very definite and unmistakable, and represented the condition desired for the stationary pattern of the received pulses.

With the time base adjusted in this manner, the circuit of figure 3 *a* was then used, and it was found that a perfectly stationary picture of the pulses was immediately obtained.

As a matter of interest the effect was then examined of connecting the time base to the successive phases of a 3-phase supply, leaving that of the transmitter fixed so as to simulate the conditions that might be met with in practice. The results were as shown in figure 3 *b*, the pulse picture moving along the base with each phase. Actually when the transmitter and time base were fed from what was known to be the same phase, the resultant picture was the last of those shown. If, therefore, in a practical system utilizing this scheme this particular picture should result, it would be easy at the receiving end to change the phase of the supply serving the time base so as to place the picture at a more suitable position, e.g. that occupied by the first picture in figure 3 *b*, in order to get the optimum conditions for observation of echoes following the reception of the ground-ray signal. In these experiments the pulses applied to the oscillograph were the radio-frequency trains derived from the coil *L* coupled to the oscillator as in figure 1 *a*. The pulses thus appear on both sides of the time base as shown in figure 3 *b*.

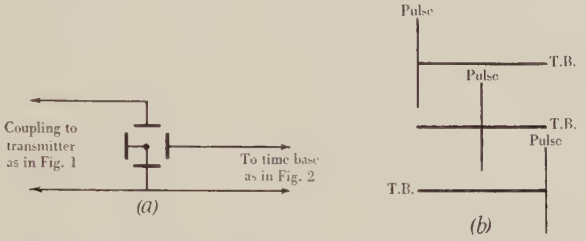


Figure 3. *a*. Circuits for checking locking of transmitter against similarly locked time base. *b*. Location of pulse as different phases of 3-phase 50 ~ supply are used for time base.

With the optimum adjustment of the time base, as in figure 2, there may be a slight tendency for successive strokes of the time base to be slightly unequal owing to inequalities of successive half-waves of the supply. This condition is very evident by visual inspection, the lowered luminosity of the excess length of one stroke being obvious. If this is observed it can be remedied by the insertion of resistance in one limb until the time base is clearly of equal luminosity over its length.

§4. USE OF UNRECTIFIED A.C.

Later exactly similar arrangements were applied in the case where both local-transmitter and receiver time-base circuits were fed with unrectified alternating current. The half-wave rectifier was omitted in each case, and the transformer was applied directly to the anode circuit of the oscillating triode, so that the valve was energized only during those half-cycles which rendered the anode positive. It was then, of course, necessary to adjust the *CR* circuit at the transmitter and the corresponding condenser-diode circuit in the receiver time base to the duration of $\frac{1}{50}$ sec., when it was possible to obtain exactly the same locking phenomena as are described above for the case of supply by rectified half-waves.

The experiments described above were all conducted on a laboratory-test scale, and appeared perfectly satisfactory. When an attempt was made subsequently to apply similar methods to transmitters of higher power, difficulties were encountered which led to modifications described later in this paper.

§5. LOCKED TIME-BASE CIRCUITS USING A MERCURY-VAPOUR TRIODE

A disadvantage of the "ticking-grid" time-base circuit described above is that, in its operation, a fairly high radio frequency is impulsively generated and, even with reasonable precautions, there is a considerable tendency for it or its harmonics to be picked up by the main radio receiver with which the circuit is, in practice, associated. Other methods of locking the oscillograph time stroke on the mains frequency are therefore desirable, and several arrangements have been tried.

A method which has proved very convenient is one using a mercury-vapour triode of the type generically known as the "thyratron." The use of a triode of this type in a self-flashing system has already been described by W. B. Nottingham* and A. L. Samuel†; the arrangement takes advantage of the difference between striking and extinction potentials, somewhat after the manner already known in connexion with self-flashing neon-lamp circuits.

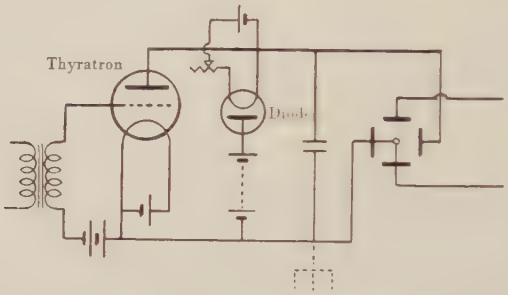


Figure 4. Use of mercury-vapour triode (thyratron) for linear time base.

A convenient modification for use as a general time base of variable frequency consists in the employment of a condenser-diode circuit as shown in figure 4. For observation of a sustained phenomenon, e.g. a steady alternating voltage, locking can be obtained by a light coupling between the grid of the thyratron and the output of the source, as indicated in figure 4. Locking in this manner has been found to give an exceedingly stable pattern, even when the frequency of the vertical phenomenon is many times—e.g. 10 or more times—that of time-base recurrence. Incidentally, it may be mentioned that the coupling necessary to give this locking effect is extremely small. For example, with a transformer of intervalve type in the grid circuit, it has been found sufficient to join only one terminal of the primary to the high-potential terminal of the source under examination. The load imposed on the source can thus be made negligibly small.

* *Journal Franklin Institute*, **211**, 751 (1931).
† *Rev. Sci. Inst.* **2**, 532 (New York, 1931).

This arrangement also permits the locking of the time stroke to the recurrence-frequency of $\frac{1}{50}$ sec., as set by 50 ~ mains. This is done by first adjusting the condenser-diode circuit to this period and then feeding into the grid coupling a small component of the 50 ~ mains voltage. The arrangement gives very satisfactory locking which, further, remains adequately locked to this recurrence rate when transients, such as the short-duration radio signals described earlier, are applied to the other dimension of deflection. A notable feature of the arrangement, true at least of the particular mercury-vapour triodes used in these experiments at low frequencies, such as those of the 50 ~ supply mains, is that the ratio of forward to back stroke can be made very short indeed, ratios well over 1000 : 1 having been found possible*.

If, however, the main purpose is to lock the receiver time base at a recurrence-rate governed entirely by the frequency of the a.c. supply, a very convenient arrangement can be obtained by feeding the anode circuit of the thyratron from the a.c. mains. A typical characteristic property of the mercury-vapour triode is that, with a suitable value of negative voltage applied to its grid, relations between anode

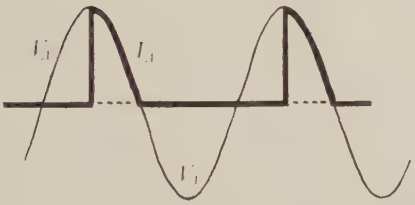


Figure 5. Abrupt rise of anode current in thyratron with suitable grid voltage.

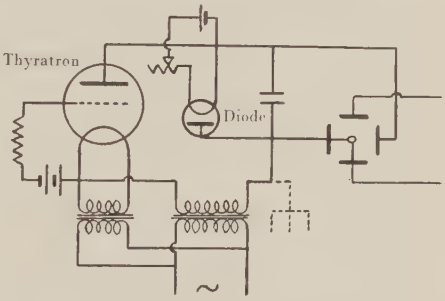


Figure 6. Thyratron circuit controlled by 50 ~ supply mains.

voltage and anode current can be made to follow a law such as those shown respectively by the dotted lines and heavy lines of figure 5. Expressed generally, according to the grid voltage actually used, it can be arranged that no anode current passes until a certain anode voltage is reached in the positive half of the applied a.c. cycle. Once anode current has thus been started, the applied anode voltage takes charge and the anode current follows the anode voltage until the latter falls to zero, when the grid voltage again takes charge.

A time-base circuit taking advantage of this principle is shown in figure 6. The negative grid voltage is set to a value such as to give conditions approximating to those of figure 5. The abrupt rise of anode current charges the condenser, which then proceeds to discharge through the diode, giving a typical time-sweep very closely approaching linearity. The condenser-diode circuit is, of course, normally adjusted to $\frac{1}{50}$ second, but the arrangement locks equally well on submultiples of the 50 ~

* Certain peculiarities of the thyratron circuit arise when the time base is caused to sweep at higher recurrence-frequencies, e.g. 1000 per second or more. These are due to the definite time taken to de-ionize and are discussed in the paper by W. B. Nottingham already cited (first reference on p. 226).

frequency. Another very valuable feature is that by shortening the discharge-period of the condenser-diode circuit—e.g. by brightening the diode filament—the time stroke can be made to remain recurrent at the rate of 50 per second, but the actual travel across the oscillograph screen can be made in a few milliseconds, with a standstill at the end of the traverse pending the incidence of the next charging impulse. Thus it can be arranged that the timing stroke shall occur in 2 or 3 milliseconds with a standstill for the remaining 18 or 17 msec., the total period being 20 msec. (i.e. 50 ~) recurrence.

This is, of course, of great advantage in giving wide opening of short-duration phenomena, provided that these can be caused to occur in the appropriate first few milliseconds of the time-sweep. To ensure this, some form of phase-control may be necessary, this being most conveniently applied to the anode-circuit voltage supply.

§ 6. A SPIRAL TIME BASE CONTROLLED AT SUPPLY FREQUENCY

While the above base may be used in the manner described to give a very rapid stroke of controlled recurrence, it has the disadvantage that the remainder of the recurrence interval is lost for observation or recording. If the whole of the phenomenon under examination can be suitably located within the brief moving sweep, high temporal discrimination can be obtained. For example, with an oscillograph screen having an effective diameter of 8 cm., a spacing of 4 cm. msec. is readily attainable. On the other hand if the whole 20 milliseconds must be kept under observation, the same effective screen-diameter only permits a spacing of less than $\frac{1}{2}$ cm./msec. In addition, if the phenomenon under observation gives an amplitude of deflection small compared with screen-dimensions, a considerable part of the screen is not fully utilized.

An obvious improvement is to turn the time scale into a circle using the same effective screen diameter; this confers immediately an improvement of π in temporal discrimination. A circuit providing such a circular time scale has been developed at the Radio Research Station and has recently been described*. The principles of the arrangement are, however, briefly recapitulated here. The circuit is as shown in figure 7. A sinusoidal voltage of frequency n is divided into quadrature components, which are applied to a push-pull modulator of the type first described by Heising. The output from the anode circuits of the valves causes the fluorescent spot to rotate in a circle whose circumference is executed in $1/n$ sec. The phenomenon under observation is introduced as shown at V in figure 7 and serves effectively to modulate the amplitude of the output, thus producing radial departures from the circular trace.

A modification giving still higher temporal discrimination, while preserving the whole of the recurrence-period under observation, consists in converting the circle into a spiral, each of whose turns is $1/n$ sec. in duration while the whole spiral base is N/n sec. in duration, where N is the number of turns in the spiral. In the case where the control is locked to the a.c. mains it is desirable to preserve the whole period N/n at 20 msec., corresponding to a 50-per-second recurrence. n thus becomes most conveniently a double or quadruple multiple of 50, secured by doubling

* *J. Inst. E. E.* **71**, 82 (1932).

or quadrupling the mains frequency, and a spiral of total length 20 msec., divided into 2 or 4 turns of 10 and 5 msec. respectively, is obtained.

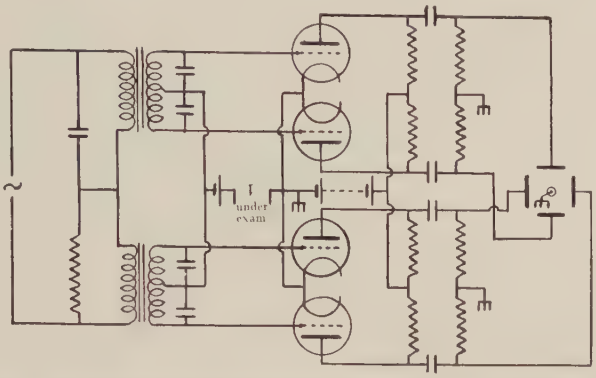


Figure 7. Circuits for circular time base with radial excursion of phenomenon under observation.

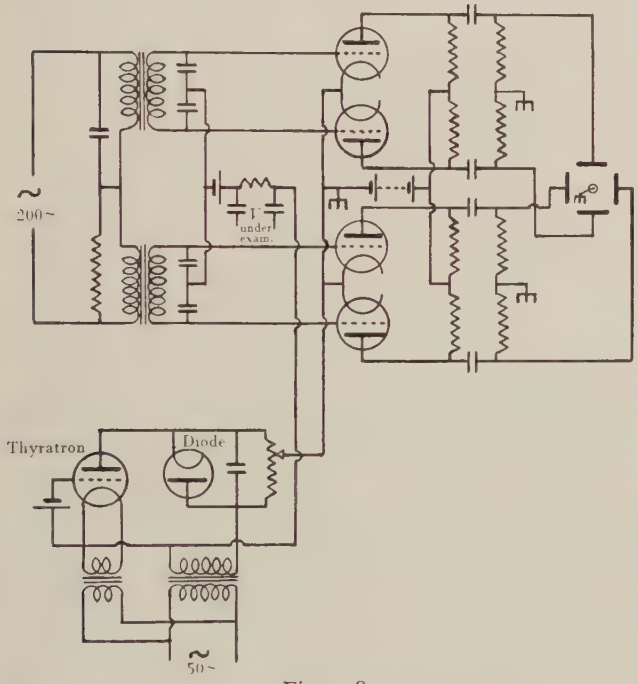


Figure 8.

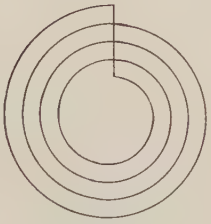


Figure 9.

Figure 8. Modification of figure 7 to give spiral time-scale.
Figure 9. Typical appearance of four-turn spiral controlled by 50 ~ mains.

The modifications necessary are shown in figure 8. Two successive stages of frequency-doubling are used to provide the input frequency to the circular-time-base system, causing the spot to rotate at the frequency of 200 ~. A saw-tooth voltage of 50 ~ is then introduced into the modulating position so that the amplitude of the circular travel is steadily diminished during four revolutions, returning abruptly to its outer position after the manner shown in figure 9. The 50 ~ saw-

tooth impulses can conveniently be obtained from a thyratron circuit controlled by the a.c. mains, as described in connection with figure 6. In order to obtain relatively uniform separation between the turns of the spiral, it is convenient to make the saw tooth linear, for instance by the use of a diode. Potentiometric division gives control of the actual amplitude of the inward movement of the spiral. If a relatively low-impedance (i.e. bright) diode is used, a high parallel resistance, used in this manner, still permits adequate approximation to linearity. The phenomenon under observation is introduced at an extra pair of terminals in the modulating position, as shown in figure 8.

It will thus be seen that a very high degree of temporal discrimination can be obtained during the first turn of the spiral, permitting close examination of the incidents of this epoch. By limiting the inward movement of the spiral, approximately as in figure 8, moderately high discrimination can still be preserved during the remainder of the incursion, thus permitting adequate examination of subsequent phenomena.

For many purposes—including, for example, the study of wireless echoes of short delay—a two-turn spiral gives sufficient improvement of time-scale length. This is, of course, readily obtained by using only a single stage of frequency-doubling, the inward travel being still controlled by the main 50 ~ frequency.

§7. AN IMPROVED METHOD OF TRANSMITTER-LOCKING

The work described at the beginning of this paper in connection with the locking of the transmitter to the mains was, as there stated, conducted only on a laboratory scale. Thus the local transmitter was simply a small power valve of the receiving class, energized by anode peak voltages of the order of 200 to 250. In these conditions no difficulty was experienced in obtaining adequate locking which remained completely stable without attention for periods of several hours and was, moreover, reproducible without further adjustment from day to day over three or four days of test observation.

In applying the same method to a larger-power transmitter using peaks of 2 or 3 kV., difficulty was experienced; it was apparently due to insulation, for example in the grid circuit. Thus, while it was possible to obtain exact locking as on the low-power scale, the adjustment was certainly more critical and the maintenance of locking less stable. For prolonged periods of observation this was sufficiently pronounced to be detrimental, at least without some attention at the transmitter, to the main advantage of the system, i.e. to the production of a perfectly stable pattern on the oscillograph screen. The ticking-grid oscillator is naturally a somewhat unstable device, and it is a common experience that, in the absence of an imposed lock, it is liable to variations of recurrence-frequency; and these variations, although possibly very small, may be sufficient to give a drifting pattern on the oscillograph. In photographic recording of the type envisaged, no drift is permissible. The imposition of a lock by the method already described effected a considerable improvement, but minor variations of insulation still permitted the tuned circuit to be triggered into

oscillation at slightly differing points about the relatively flat top of the a.c. positive peak-voltage.

In order to improve the conditions, it was thought jointly by Mr G. Builder and the author that advantage might be taken of the property of the thyratron already illustrated in figure 5 and described in connection with the locked time base. In particular, it appeared probable that the abrupt rise of anode current in the thyratron circuit could be made to act as an additional switch or master control, triggering the circuit into oscillation very abruptly at a definite point of the anode-voltage cycle. A thyratron was therefore inserted in series with the anode circuit of the valve, as shown in figure 10. Uncertainty existed as to the exact values of the

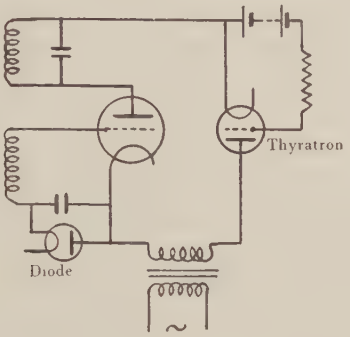


Figure 10. Control of pulse-generation by thyratron as abrupt switching-device.

voltage-drop applied at any instant across the thyratron and valve respectively, especially during their inactive periods, and this rendered it difficult to make any pre-calculation of the voltage necessary to give the appropriate negative grid bias to the thyratron. When a local oscillograph, however, was used as a monitor in the manner already described, no practical difficulty was experienced in obtaining a value of bias which materially improved the locking state, and yielded also another desirable improvement, viz. a shortening of the signal pulse. The value of the bias was not found to be highly critical in adjustment, while the addition of the thyratron was found to confer much greater latitude in the other adjustments, notably those of the condenser-diode combination which was used in the grid circuit as a convenient means of control.

§ 8. CONCLUSION

The locking of the transmitter and the oscillograph image represent a matter of considerable importance in any prolonged series of observations using this type of signal for measurements of the ionosphere. While the arrangements discussed by no means exhaust the possibilities of the use of common-frequency supply mains at both stations, they represent methods which are economical in execution and satisfactory in operation. Although they have been described in what is no doubt a

particular and perhaps a somewhat limited application, the methods may possibly have other applications in the examination of various types of phenomena, and particularly of transients, where the recurrence of the time base and of the phenomenon under observation can be controlled from a common-frequency source.

§ 9. ACKNOWLEDGMENTS

The work described in this paper was carried out as part of the programme of the Radio Research Board and is published by permission of the Department of Scientific and Industrial Research. Thanks are due to Mr R. A. Watson Watt, Superintendent, Radio Research Station, Slough, for the initial suggestions in connection with this work and for continued advice during its progress.

DISCUSSION

Prof. E. V. Appleton. The author's contribution to our technique for ionospheric investigations constitutes an important step forward. By putting the burden of synchronization on the employment of a common electric mains supply at sending and receiving station we not only render visual observations of echo portrayals easier but also bring about the possibility of continuous photographic recording. Apparatus designed on the lines described by Mr Herd and built at the Radio Research Station, Slough, was taken out last July to Tromso in Norway in connexion with the work to be carried out during the International Polar Year, 1932–3. Owing to the fact that the same alternating-current supply was available at Simavik (on the island of Ringvadsø) and Tromso, which places are 19 km. apart, it was possible to use the author's method of synchronization. The fact that the expedition was able to build its sending and receiving stations on the two islands in question and begin recording within a week of landing at Tromso is a tribute to the simplicity and reliability of the system.

Mr R. A. Watson Watt. I think the papers by Messrs Herd and Naismith might appropriately be described as "Workshop Notes." They represent two aspects of the forging and testing of appropriate tools for the investigation of the ionosphere. This general investigation is not a "wireless" one in the sense that it is of interest only to the communications engineer. Ionospheric research is an important branch of the general physics of the earth, of that wider "geophysics" whose title is tending to be, somewhat unfortunately, usurped by the applications of a small branch of the subject.

Technical tricks, such as those described by the author, are of extreme importance where it is necessary to amass large quantities of observational material; the reduction of the material is so much more laborious than is the mere recording that every effort must be made to carry out recording in the form which makes

reduction easy and expeditious. The methods here described are being applied very successfully to this end in the special Polar Year work which the Department of Scientific and Industrial Research, with the co-operation of the National Committee for the Polar Year and of King's College, London, is carrying out within the Arctic Circle and in this country. But I have no doubt that the very neat and effective devices described by the author will find applications in other fields, and not only in those where "wireless" problems are in question.

The author has perhaps erred on the side of modesty in failing to indicate that he was very directly concerned in the second reference which he gives at the foot of p. 221.

Mr P. W. Willans. I was interested in a remark made by the author on the inequality of the time-base strokes in one of his synchronizing arrangements. He attributes this inequality to the fact that his full-wave rectifying device was not operating with strict symmetry. My experience has been that a similar effect may be produced as a result of the nature of the impedance in the anode circuit of the "ticking oscillator" valve. In the course of some attempts to make a ticking oscillator produce pulse wave-forms of various shapes in its anode circuit I more or less accidentally inserted a tuned circuit of high efficiency in this circuit, the resonant frequency being some six or seven times the low-frequency characteristic of the ticking oscillator. A somewhat astonishing result was obtained, in that two alternating trains of dissimilar damped waves were set up in the tuned circuit, the effect being readily observable on a cathode-ray oscillograph. One train started from a larger amplitude than the other and the impulse at the end of the first train gave rise to a sudden alteration of the amplitude of such a character as to set up a second train of smaller amplitude which in its turn was followed by an exact repetition of the first train, so that a steady succession of alternating trains was produced. No opportunity has presented itself of following this matter up, but the possibility of such a phenomenon producing an effect of frequency sub-multiplication constitutes an argument of caution as regards the use of a form of control operative in the anode circuit of a ticking oscillator, unless the nature of the impedance is such that the effect mentioned cannot take place.

Author's reply. There is no doubt that such methods of synchronization have proved of high importance in simplifying the observational technique of ionospheric study, and we are still endeavouring to effect further refinements in the way of simplifying routine work and assisting both the annotation and reduction of the records.

With reference to the matter raised by Mr Willans, I can visualize the possibility of the action he suggests, especially in the case of the circuit which I know he was using. I feel sure, however, that the particular effect reported in the paper was definitely due to inequality of the rectified half-waves. I proved this to my complete satisfaction at the time by the use of resistance with which it was possible gradually to reduce and balance the inequality, and then, by continuing the addition of

resistance, to introduce inequality in the opposite direction. It was possible, of course, to check this process in its various stages, while it was significant that the effect disappeared completely when, instead of rectified half-waves from a full-wave rectifier, unrectified a.c. was used, the valve operating on the positive half-waves as described later in the paper.

I should apologize for a drawing-error in figure 3 (*b*), which does not correctly show the 120° displacement due to successive phases of a 3-phase supply. It is, however, indicative of the manner in which the received pulse can be located at a suitable position along the time-stroke.

551.594.11

A COMPARISON OF THE FREQUENCY-CHANGE AND GROUP-RETARDATION METHODS OF MEASURING IONIZED-LAYER EQUIVALENT HEIGHTS

By R. NAISMITH, A.M.I.E.E., Radio Research Station, Slough

Communicated by Prof. E. V. Appleton, F.R.S., September 20, 1932.
Read December 2, 1932.

ABSTRACT. The paper describes a number of experiments designed to test the measurement of the equivalent heights of the ionized layers by the frequency-change and group-retardation methods. Measurements are conducted under as varied conditions as possible. The results appear to confirm the theoretical investigation recorded by Appleton in 1928. The apparatus used to measure the amount of the frequency change employed in the measurements is described.

§ 1. INTRODUCTION

Within the last decade a number of different methods have been proposed to determine the height of the ionized regions of the upper atmosphere. Some of these methods have proved more successful than others, with the result that in the more recent work on the study of these regions the two methods which form the title of this paper have been most widely used. The first, known as the frequency-change method, was used by Appleton and Barnett* in their original experimental proof of the existence of the ionized layer and has formed the basis of much later work in this country. With balanced antenna systems this method has been used successfully to study polarization and other phenomena, which up to the present has not been possible by the group-retardation method. In the application of this first method, the frequency of an emitting station is altered continuously through a small range and the interference maxima are noted at a distant receiving station. Such interference takes place between the waves which have travelled along the ground and those which have travelled through the upper atmosphere. In the second method, known as the group-retardation method, and due to Breit and Tuve†, the emitting station sends out short pulses of radio-frequency energy and the time interval is measured between the arrival of the pulse via the ground and that via the atmosphere. At first sight it would appear that these two types of measurement are not concerned with the same quantities, since in one case interference phenomena are in question and in the other group-time measurements. Until quite recently this possible difference was emphasized by the fact that widely different interpretations had been placed on the results obtained by the frequency-change method in this country and by American workers using the group-retardation method.

* *Nature*, p. 333 (March 1925) and *Proc. R. S.* A, 109, 621 (1925).
† *Phys. Rev.* 28, 568 (1926).

§ 2. OBJECT OF EXPERIMENT

In 1928, however, it was shown theoretically by Appleton* that, for the circumstances under which the experiments are at present carried out in this country, the two methods should give the same result. This equivalence can be illustrated in the following way.

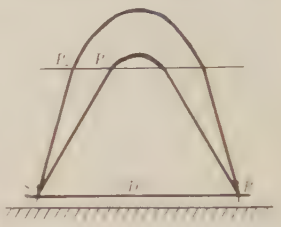


Figure 1.

D Consider a case of transmission from S to R. Let $SR = D$

P_1 and let P_1 be the optical path of the wave-track of the

f_1, P_2 atmospheric waves when the frequency is f_1 while P_2 is

f_2 the corresponding wave track when the frequency is f_2 (note that P_1 and P_2 are not identical if the refracting medium, as in this case, is dispersive). If interference fringes

n are produced at R there will be experienced n signal maxima (or minima) if the frequency is changed continuously from f_1 to f_2, where

$$n = \frac{1}{c}\{(P_1 - D) f_1 - (P_2 - D) f_2\} \qquad \ldots\ldots(1)$$

c and c is the velocity of radiation in free space. This formula is valid when the change of frequency $(f_1 - f_2)$ is large. Under the conditions of the experiment, however, the change of frequency is small and we can write in differential notation

$$\frac{\Delta n}{\Delta f} = \frac{1}{c}\left\{(P - D) + f\frac{dP}{df}\right\} \qquad \ldots\ldots(2),$$

or

$$\left(c\frac{\Delta n}{\Delta f} + D\right) = P + f\frac{dP}{df} \qquad \ldots\ldots(3),$$

P where P now refers to the optical path of the mean wave-length used. The quantity $(P + f\,dP/df)$ Appleton calls the "equivalent path" of the atmospheric waves.

t In the group-retardation method it is clear that, if t is the difference in the times of arrival of a certain pulse via the ground and via the upper atmosphere, we have

$$t = \int\frac{ds}{U} - \frac{D}{c} \qquad \ldots\ldots(4),$$

where U is the group velocity of the atmospheric waves along an element of atmo-

ds spheric ray track ds and the integration extends along the whole length of this track. We then have

$$ct + D = C\int\frac{ds}{U} \qquad \ldots\ldots(5).$$

Now Appleton, using the principle of least time as interpreted by Hamilton, has demonstrated the equivalence of the right-hand sides of (3) and (5) so that the quantity $\Delta n/\Delta f$ measured in the frequency-change method should be equal to t measured by the group-retardation method. If these quantities are shown to be equal, the equivalent paths of the atmospheric rays and therefore the equivalent heights of the deviating layer deduced from them should be the same.

* *Proc. Phys. Soc.* **41**, 43 (1928).

The object of the experiments described below was to see how far this suggested theoretical equivalence was borne out in practice when both methods were carried out under identical conditions.

§ 3. EXPERIMENTAL PROCEDURE

In the frequency-change method it is necessary to know both the amount of variation of frequency and the number of maxima and minima caused by that change of frequency. It is desirable that both these determinations should be made at the receiving station. The scheme now adopted is to set the amount of frequency change to approximately 16 kc./sec. and to measure the change accurately by a method which is described at length at the end of this paper. The signal maxima and minima were recorded on an Einthoven galvanometer with camera attachment. Preliminary experiments with a special form of syphon recorder show that it will eventually be possible to obtain these variations in signal strength directly on a tape and without the necessity for photographic equipment. With these modifications, the apparatus necessary for the experimental determination of the height of either the E or the F ionized regions of the upper atmosphere by the frequency-change method is now comparatively simple, robust and direct reading.

The pulses required for the group-retardation experiment were produced similarly to the manner described by Appleton and Builder in January 1932*. These pulses were observed on a cathode-ray oscillograph after suitable amplification. When it was desired to obtain a record of any observed pattern the photographic recording was done on a Duddell high-frequency oscillograph. One element of this oscillograph was coupled to a 1 kc./sec. oscillator which could readily be calibrated against a 1 kc./sec. tuning-fork. A second element of the oscillograph recorded the pulses due to the signal coming over the ground and the echoes present. In this way it was possible to compare directly the time interval between the initial pulse and the succeeding echoes with a 1 kc./sec. oscillator, and thus obtain the path difference between the pulse which had travelled by the upper regions of the atmosphere and that which had travelled directly over the surface of the ground from the transmitter to the receiver.

In the individual experiments quoted below a standard system of transmission was adopted. First of all a determination of the effective height of the ionized layer was made by the frequency-change method. This was followed immediately by a determination by the group-retardation method and finally by another determination by the frequency-change method. In this way it was possible to eliminate any possibility of sudden changes in the effective height at which reflection took place. The transmitter, except where the contrary is stated, was situated at the National Physical Laboratory at Teddington, and the receiver at Slough. This represents a ground distance of 18 km. The accuracy of the frequency-change measurement is estimated to be within 1 per cent, but on the group-retardation method it may not be greater than 2 per cent or even 3 per cent where measurements are made on the lower ionized region.

* *Proc. Phys. Soc.* 44, 76 (1932).

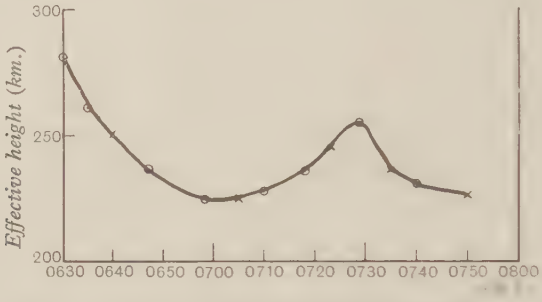
§4. EXPERIMENTAL RESULTS

The first experiment was a straightforward comparison on 85 m. It was repeated at intervals throughout the winter and table 1 gives a summary of the results obtained.

Table 1

Date	G.m.t.	Height (km.) measured by	
		Frequency change	Group retardation
30. x. 1931	12.30	253	251
3. xii. 1931	12.30	218	224
25. ii. 1932	12.00	220·5	223
7. iv. 1932	12.30	235	233

Many more observations have been made on this wave-length, but these results still represent typical values. In no case was the discrepancy greater than 2 per cent.



× Heights measured by group-retardation method
⊙ „ „ „ frequency-change „ } 26. 2. 32.

Figure 2.

The second experiment was made on February 26, 1932, over the sunrise period in order that comparisons might be made at a time when the effective height of the ionized region was changing steadily. These results are best shown in the form of a graph (see figure 2). The same wave-length of 150 m. was used throughout the test. At 08.00 G.m.t. it was not possible to make a determination of the height owing to the fact that the density in the lower layer had increased sufficiently to cause this wave-length to be about the critical one. At 08.30 this wave-length ceased to penetrate the lower ionized region and was reflected from a height of 119 km.

The third experiment was arranged to make a comparison of the methods using the lower layer as a basis. The wave-length chosen was 150 m. and table 2 gives an example of the results obtained.

Table 2

Date	G.m.t.	Height (km.) measured by	
		Frequency change	Group retardation
25. ii. 1932	11.15	131	128
11. iii. 1932	9.15	107	108
7. iv. 1932	11.15	108	108

On April 14, 1932, the density of the *E* layer was abnormally high. This provided an opportunity to make a comparison of the methods on a very much shorter wave-length than had hitherto been possible with the *E* layer as the basis of comparison.

On 85 m. the heights measured at 12.45 G.m.t. were as follows:

Frequency change	101·5 km.
Group retardation	102 km.

An experiment on January 8, 1932, between Windsor and Slough over a ground distance of 5 km. determined the lowest wave-length which was reflected from the *F* layer. The result of this experiment showed that radiation of wave-length 45 m. was reflected but that of wave-length 42·5 m. was not. A comparison was then made on 45 m. and the heights measured at noon were as follows:

Frequency change	223 km.
Group retardation	226 km.

Another arrangement was tried with different transmitters and over different distances. The wave-lengths were made slightly different (90 and 88·3 m.) in order that practically simultaneous observations might be made. In one case the transmitter was installed near Windsor which is 9 km. from the receiver at Slough, and in the other at Rugby which is 111 km. distant. The effective height measured from Windsor was 211 km.; this was obtained by the frequency-change method. The path difference as determined by the Rugby transmission was 422 km., representing an effective height of 214 km. This latter was determined by the group-retardation method.

The large change in the angle of incidence on the layer in the two cases appears to have produced very little change in the depth of penetration into the layer.

§ 5. CONCLUSIONS

The above typical set of experiments were designed as far as possible to test the comparison of the frequency-change and group-retardation methods of measuring the effective height of the ionized regions of the upper atmosphere under a variety of conditions. Other observations which have been made at different times only serve to confirm in experimental terms the theoretical conclusion arrived at by Appleton in 1928 that the frequency-change method and the group-retardation method both measure the same quantity, namely,

$$c \int \frac{ds}{U}$$

for the atmospheric waves.

§ 6. ACKNOWLEDGMENTS

The work described in this paper was carried out as part of the programme of the Radio Research Board and is published by permission of the Department of Scientific and Industrial Research. Thanks are due to Professor E. V. Appleton, F.R.S., for initial suggestions in connection with this work. Thanks are also due to

Mr R. A. Watson Watt, Superintendent, for his interest in the work and for the provision of facilities for its conduct, and to other colleagues at the Radio Research Station, Slough, for assistance in the experiments.

APPENDIX

A device for the measurement of a small change of frequency

Object. The object of the device is to provide a means of measuring the small change of frequency required for the experimental determination of layer heights by the frequency-change method.

Principle. The principle of the device is to change the frequency f of the received signal to another frequency f_1 such that the amount Δf (by which f is changed) to be measured is sufficiently large compared with f_1 to be read directly from an oscillator calibrated over the range Δf and centred on frequency f_1.

Description. Figure 3 is a simple schematic arrangement of the device which is almost self-explanatory. In practice the first detector is tuned to the received signal. The wave-meter is tuned 100 kc. sec. off, providing a beat frequency of 100 kc. sec. which is applied to the grid of the second detector along with the output from the

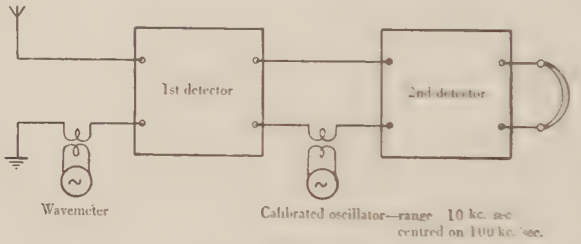


Figure 3.

calibrated 100 kc./sec. oscillator. This gives zero beat after the second detector. When the frequency changes by the amount Δf, zero beat can again be obtained by retuning the 100 kc./sec. oscillator. The amount by which this oscillator has been retuned is then read off directly in kilocycles.

Certain modifications of a very practical nature can be incorporated when the particular conditions of the experiment are known. For example when it is only desired to check a frequency change known to be approximately 16 kc. sec. a condenser can be connected by a switch across the tuned circuit of the 100 kc./sec. oscillator which will change the frequency by that amount.

All that is then necessary is to tune to zero beat and when the frequency starts changing throw the switch, and when the change of frequency stops zero beat should again be heard. If this is not so, a readjustment of the variable condenser will produce the condition and the amount of this readjustment (+ or −) will give the departure from 16 kc. sec. Other modifications (e.g. first detector oscillating and no wave-meter) can be readily incorporated but these need not be discussed here.

Advantages. No local signal of the same frequency as the received signal is present. The measurement and the photographing can therefore be done simultaneously and each individual frequency change can be measured. The time taken for the measurement, after the apparatus has been tuned to the signal, can be made as short as 3 sec. when the amount of the frequency change is known approximately. Automatic synchronization between oscillators or oscillator and transmitter is entirely absent.

The following critical test was applied. A very weak c.w. signal of 75 m. was used. The apparatus was tuned to this wavelength and the intermediate frequency oscillator adjusted until a beat note of 1 kc./sec. was heard. A low-frequency oscillator was tuned to 1 kc./sec. in the same room and the audible beats between the oscillator and incoming signal were detected by the ear. A very powerful local signal was then produced and tuned through 75 m. This signal as it passed through the tune of the high-frequency circuits did not cause the beats set up by the weak signal and the 1 kc./sec. oscillator to vary. It is considered that this test is sufficiently stringent to show that the coupling between the amplifier tuned to 100 kc./sec. and the 100 kc./sec. oscillator is sufficiently weak to eliminate any error due to this cause. The accuracy is easily within 1 per cent. A difference method is employed. This has some important advantages over systems referred to absolute standards. The most important is perhaps the simplicity of the scheme. Also any small drift from the frequency to which the oscillators are tuned is relatively unimportant. There can hardly be any ambiguity caused by beating with a harmonic instead of the fundamental.

It is possible to check whether the frequency is increasing or decreasing during any change, thus: The h.f. oscillator is tuned above the frequency of the incoming signal and the i.f. oscillator is tuned about 2 kc./sec. above the frequency to which the i.f. amplifier is tuned. The note will then rise or fall during the frequency change according to whether the change is due to an increase in the frequency of the incoming signal or to a decrease. Similar results are obtained when both oscillators are tuned below the respective frequencies.

DISCUSSION

Prof. E. V. Appleton. A careful comparison of results by the two available methods, such as the author has made, was very much needed and it is satisfactory to note that such good agreement has been obtained. Since we are concerned essentially with the time required for a signal to travel up to the ionosphere and back we have, as it were, to put a mark on the wave by which to recognize it. Now a wave has only two characteristics which we can control, namely frequency and amplitude. In the frequency-change method we use frequency-modulation and in the group-retardation method we use amplitude-modulation. Any other method we can devise must be a modification of the two basic ones.

Mr R. A. Watson Watt. I have already commented on the broad aspects of the author's paper in discussing the paper by Mr Herd. The good agreement shown by the experiments is, as Prof. Appleton has remarked, very satisfactory. I think the report of the experiment in which emissions from Windsor and Rugby respectively formed the material for comparison is of very special interest; it is one of many indicators towards an extended programme of comparisons between vertical-incidence phenomena on one hand and inclined-incidence phenomena on the other. It might be of interest for Mr Naismith to record whether the Rugby determination was made on ground ray and first reflection or on multiple reflections.

Mr Naismith's familiarity with the state of the general investigation may have led him to overlook the fact that two *obiter dicta* of the paper, true at the time of drafting, were no longer true at the date of communication. The group-retardation technique has for some time been used in polarization studies*, and the introduction of the syphon recorder into frequency-change working, to which we were led by a suggestion from the late Captain Lemon, R.S., was so satisfactory that the syphon recorder has been systematically used at Slough since July 1932.

Dr F. J. W. Whipple. In the introduction to this paper the author follows Prof. Appleton in considering the actual wave-tracks, and the reader who is unfamiliar with the subject may, like myself, not realize at once that the "optical path P" is merely a technical name for the product of the time of transmission of a phase and the normal velocity of light. It seems better to introduce a symbol T for the time of transmission. Then we can write down the number of waves in the path, fT, and the equation (1) of the paper becomes

$$n=(f_1T_1-f_1D/c)-(f_2T_2-f_2D/c),$$

so that (3) takes the form

$$\frac{\Delta n}{\Delta f}+\frac{D}{c}=\frac{d}{df}(fT).$$

Prof. Appleton's demonstration that the right-hand side of this equation is equal to $\mathfrak{T}$, the time of transmission of a group, is really independent of any knowledge of the wave-tracks. We can bring this out by considering a pulse of a definite form, the initial impulse being proportional to $e^{-nt^2}\sin 2\pi f_0 t$.

By Fourier's theorem we find that

$$e^{-nt^2}\sin 2\pi f_0 t = 4\sqrt{\frac{\pi}{n}}\cdot\int_{-\infty}^{\infty} e^{-4\pi^2(f_0-f)^2/n}\sin 2\pi ft\,df,$$

so that the pulse may be regarded as the combination of trains of waves, those for which f is near to f_0 predominating.

The disturbance which reaches the receiving station is

$$4\sqrt{\frac{\pi}{n}}\cdot\int_{-\infty}^{\infty} A(f)e^{-4\pi^2(f_0-f)^2/n}\sin 2\pi f(t-T)\,df,$$

the factor $A(f)$ indicating that the attenuation of the waves may depend on frequency.

* *Wireless World*, July 8, 1932 and *Nature*, September 17, 1932.

The disturbance is equivalent to

$$4\sqrt{\frac{\pi}{n}}.\sin 2\pi f_0(t-T_0)\int_{-\infty}^{\infty} A(f)e^{-4\pi^2(f_0-f)^2/n}\cos 2\pi\{(f_0-f)t-(f_0T_0-fT)\}\,df,$$

a similar expression with the sin and cos interchanged being neglected. The latter expression would vanish if $A(f)$ were constant.

The disturbance is a maximum when

$$(f_0-f)t = f_0T_0-fT$$

for values of f near to f_0.

Accordingly the "group" arrives at the time t given by

$$t = \mathfrak{T} = \frac{d}{df}(fT), \quad f=f_0.$$

Obviously the result would have been the same if a different type of initial pulse had been assumed.

The author's experiments provide a beautiful verification of the identity

$$[\mathfrak{T}-D/c] = \Delta n/\Delta f.$$

It may be of interest to point out that the ordinary formula for group velocity can be deduced from the equation

$$\mathfrak{T} = \frac{d}{df}(fT),$$

by reversing Appleton's argument.

Writing U for the group velocity and V for the pulse velocity, we have

$$\int\frac{ds}{U} = \frac{d}{df}\left[f\int\frac{ds}{V}\right].$$

Now Hamilton's principle tells us that $\int ds/V$ with f constant is unaltered by slight alterations in the path of integration and therefore by the change from the path appropriate to f to that appropriate to $f+df$. Accordingly

$$\int\frac{ds}{U} = \int\frac{d}{df}\left(\frac{f}{V}\right)ds.$$

Since "group velocity" is a property of a particular set of waves and is difficult to define precisely, we cannot *prove* the formula

$$\frac{1}{U} = \frac{d}{df}\left(\frac{f}{V}\right).$$

We can say, however, that this formula is consistent with the preceding equation, and that is sufficient to justify the use of the concept, "group velocity."

Author's reply. I am indebted to Prof. Appleton for his concise remarks. He has put the essential difference between the two methods very explicitly.

In reply to Mr Watson Watt: The comparison between the emissions from Windsor and Rugby was really done as part of another more general experiment in which I was engaged, namely the comparison of measurements on the ionosphere made over different base lines. These measurements had shown that it was possible

to obtain, under certain conditions, quite different results when the base line between emitter and receiver was altered. On this particular occasion, however, the density of the *E* region was constant throughout the middle of the day when the measurements were made. The determination was facilitated by the reception of a ground pulse, and the fact that the frequency on which the comparison was made was not near the critical penetration frequency at the time.

Dr Whipple's contribution is helpful in that it presents the theoretical aspect of the subject from a different angle. This will no doubt assist certain readers to a better understanding of the problem.

621.396.11

THE APPLETON-HARTREE FORMULA AND DISPERSION CURVES FOR THE PROPAGATION OF ELECTROMAGNETIC WAVES THROUGH AN IONIZED MEDIUM IN THE PRESENCE OF AN EXTERNAL MAGNETIC FIELD

PART 1: CURVES FOR ZERO ABSORPTION

By MARY TAYLOR, M.A., Ph.D., Radio Research Station, Slough

Communicated by Mr R. A. Watson Watt, November 21, 1932.
Read December 16, 1932

ABSTRACT. This paper gives dispersion curves derived from the Appleton-Hartree formula in the case of zero absorption. The value of the magnetic field is taken as that of the earth's field at Slough. The curves are drawn to show the value of the squares of the indices of refraction and attenuation as functions of the electron density for a series of twelve frequencies, which are chosen to illustrate the various classes of curve and the boundary curves separating the classes and, in the case of frequencies above 1·321 megacycles per second, the various regions of short and ultra-short waves. The derivation and general properties of the Appleton-Hartree formula and the various possible modes of propagation are also discussed. The dispersion curves are classified according to the infinities they contain and a diagram is given to show how the classes of curve holding for any angle of inclination of the direction of propagation to the magnetic field H depend on the ratio of the longitudinal component of H to H itself. The use of the zeros and infinities of the dispersion curves in the interpretation of propagation phenomena is described and a summarizing diagram is given, showing how the possible propagation of zero, one or two basic modes for any frequency depends on the electron density. The polarization corresponding to each dispersion curve is shown graphically and the general properties of the polarizations of the basic propagation modes are discussed.

§ 1. INTRODUCTION

It is now well established that the presence of a magnetic field exerts considerable influence on the propagation of electromagnetic waves through a dispersive medium. The magnetic rotation of the plane of polarization was discovered by Faraday in 1845, and the Kerr and Zeeman effects in 1877 and 1896 respectively. It is to Lorentz that we chiefly owe a theoretical explanation of the phenomena on classical lines, in contradistinction to the methods of modern quantum mechanics. He assumed that the dispersion phenomena were due to the presence in the dispersive medium of discrete charged particles. An expression for the electric force acting on any one particle was found by combining the average effect of the other particles with the electric vector of the incident wave; this was used in the equations of motion of the particle (which included the effects of the external magnetic field) to get a relation between the vectorial electric force E and the vectorial dielectric

E

D displacement D. Lorentz combined this relation with Maxwell's electromagnetic equations in the two simple cases of propagation (*a*) along, and (*b*) at right angles to the lines of the external magnetic force. He assumed that the external magnetic field acted along the positive direction of the axis Oz; in case (*a*) Oz was also the direction of propagation of the waves through the medium, and in case (*b*) they were propagated along Ox. Lorentz showed that in each case there were two values of the P_x, P_y ratio of the components P_x and P_y of the vectorial polarization P, or $(D - E)$, which governed propagation through the medium*. Each of these values corresponded to a mode of propagation through the medium in which the form of the wave or vibration was unchanged, except in so far as exponential decrease of amplitude was involved; we may denote such modes by the term *basic modes*, since any other mode may be regarded as combinations of them in varying degrees. The refractive index, index of attenuation, and polarization are fixed for each mode by the value of the real or complex ratio $P_x : P_y$.

Lorentz showed that in case (*a*), when propagation was along the lines of the external magnetic force, each value of $P_x : P_y$ corresponded to a beam of circularly polarized waves, one value giving a left-handed and the other a right-handed sense of rotation of the electric vector around the positive direction of propagation of the beam. In case (*b*), with propagation at right angles to the external magnetic force, the magnetic field was found to have no influence on the electric vibrations along the lines of force, whereas waves in which the electric vector was perpendicular to the external magnetic field were in general elliptically polarized in the plane perpendicular to the external magnetic force, though they appear as linearly polarized when viewed along the direction of propagation.

The application of this theory to the explanation of phenomena in the propagation of wireless waves through the earth's atmosphere was first proposed by Appleton in 1925(a) to account for directional errors. He drew attention to the fact that the quantity He/mc, the natural angular frequency of rotation of electrons about the earth's magnetic field, was comparable with wireless frequencies and so must be taken into account in theories of propagation. He suggested that the presence of an abnormally polarized wave giving directional errors might be due to the joint action of the ionized medium and the earth's magnetic field, which must obviously in principle affect the propagation of the waves. Since, however, the direction of propagation of the waves is seldom either along or perpendicular to the direction of the lines of force, it was necessary to extend Lorentz's treatment to the case when the magnetic field is inclined at any angle to the direction of propagation of the waves, and the formula determining the characteristics of the basic modes in this case was published by Appleton in 1927(c). He assumed the charged centres to be free electrons, where Lorentz had regarded them as bound. The expression given by Appleton has the form

$$M^2 = \left(\mu - \frac{i\kappa c}{p}\right)^2 = 1 + \frac{2}{2(\alpha + i\beta) - \frac{{\gamma_T}^2}{1 + \alpha + i\beta} \pm \sqrt{\left(\frac{{\gamma_T}^4}{(1 + \alpha + i\beta)^2} + 4{\gamma_L}^2\right)}} \quad \ldots\ldots(1).$$

* The Heaviside system of units, used by Lorentz, is here retained.

In this paper

c is the velocity of electromagnetic waves in vacuo;	c
m the mass of an electron;	m
e the charge on an electron (e.s.u.);	e
N the electron-density;	N
H the imposed magnetic field (gauss);	H
H_L the component of magnetic field along direction of propagation;	H_L
H_T the component of magnetic field perpendicular to direction of propagation;	H_T
$p = 2\pi$ times the frequency of the wave considered;	p
$p_0^2 = 4\pi Ne^2/m$;	p_0^2
$p_H = He/mc$, and is the natural angular frequency of rotation of electrons about H;	p_H
$p_L = H_L e/mc$, and is the natural angular frequency of rotation of electrons about H_L;	p_L
$p_T = H_T e/mc$, and is the natural angular frequency of rotation of electrons about H_T;	p_T
$\nu = 1/\tau$, and is the frequency of collision of electron with molecules;	ν
$\alpha = -p^2/p_0^2 - \frac{1}{3}$;	α
$\beta = p\nu/p_0^2$;	β
$\gamma = pp_H/p_0^2$;	γ
$\gamma_T = pp_T/p_0^2$;	γ_T
$\gamma_L = pp_L/p_0^2$;	γ_L
μ is the index of refraction and $= c/$(phase velocity of propagation);	μ
κ is the index of attenuation; and	κ
$M = \mu - i\kappa c/p$.	M

The same expression, with different notation, was independently derived by Hartree (3) in his study of the general differential equations of propagation in any stratified ionized medium; he summed the waves scattered by the discrete charged centres and superposed the result on the incident plane wave. The formula (1) may therefore appropriately be called the Appleton-Hartree formula. The present paper is concerned with the dispersion curves to be derived from it and used in the interpretation of wireless reception phenomena. We may note at once that, when the external magnetic field has a non-zero value, the two values of M are only equal if (1) $N = 0$ or (2) $1 + \alpha = 0$ and $\gamma_T^4 - 4\beta^2\gamma_L^2 = 0$ simultaneously; these cases are the only ones in which the two basic modes of propagation given by the alternative signs before the square root can be the same. The theory strictly holds only when the number of electrons in a cube whose side is a vacuum wave-length is large; it is thus not to be expected that the formula will correspond rigidly to physical reality for very small values of N, but it does give $M^2 = 1$ when $N = 0$, as physical considerations demand. This point arises again later, in connexion with the dispersion curves.

In deriving the formula, absorption was regarded as caused by collision of the free electrons with the neutral molecules of the gaseous atmosphere around them. The frequency ν of these collisions per electron is a variable quantity and increases

with the density of the atmosphere. The curves given in the present paper are restricted to the case $\nu = 0$; those for small values of absorption differ only slightly from those given here, though the presence of even a small amount of absorption removes the infinities on the dispersion curves, as is usual in physical theory, and also requires that μ and κ in (1) be never actually zero.

The discussion given by Goldstein[4] treated the medium as a homogeneously charged fluid, in contrast with the discrete charged centres assumed by Lorentz, Appleton and Hartree, and disregarded the effect of collisional friction which was included by the other writers named. The dispersion curves derived on this assumption may differ considerably from those given by the Appleton-Hartree formula, particularly at low frequencies.

§ 2. MODES OF PROPAGATION

In discussing the general properties of the Appleton-Hartree formula and the types of dispersion curves it gives, it is convenient to start from the form (1) in which it was derived by Appleton. The alternative signs to the square root show that M^2 has two values for any one set of values of p, N, H_L, H_T and ν, so that, generally speaking, there are two basic modes of propagation of electromagnetic vibrations of angular frequency p through a medium of constant electron-density with a given magnetic field and collision-frequency. These modes are defined and distinguished by (*a*) the two values of M^2, and (*b*) the polarizations corresponding thereto, which have been shown by Appleton and Hartree to be given by

$$-\frac{E_x}{E_y}=\frac{H_y}{H_x}=\frac{i\gamma_L(1-M^2)}{1+(\alpha+i\beta)(1-M^2)} \qquad \text{.....(2),}$$

where the incident wave is propagated along the positive direction of Oz and the external magnetic field has components H_T and H_L along Oz and Ox respectively.

The vibration actually occurring in the medium is determined for continuous waves by the proportion in which the modes are present at any time and may be found by the usual method of building up a wave in the medium by superposition of the possible basic vibrations. We suppose the incident wave to be travelling along the positive direction of Oz. If we distinguish the two values of M^2 by the suffixes
a, b a and b when the positive and negative signs respectively are prefixed to the square root in the denominator of the right-hand side of (1), we see, following Lorentz*, that the components of the vectors describing the electromagnetic field are represented by the real parts of expressions built up by adding the four quantities $e^{-\kappa_r z + ip(t - \mu_r z/c)}$ $(r = a, b)$, each multiplied by a constant factor determining its amplitude and phase. The factors are found in any actual case by connecting the total vibration to the incident wave of given polarization and amplitude by means of the boundary conditions to be satisfied at the surface of the ionized medium.

The phase velocity of propagation of a basic mode is c/μ_r, so that the index of refraction
μ_r $\mu_r = c/(\text{corresponding velocity of phase propagation}).$

* *Theory of Electrons*, p. 153 (1916).

The quantities κ_r determine the rate at which the amplitudes of the vibrations decrease and are called by Lorentz the indices of absorption, although, as will be seen below, true absorption of energy only occurs when $\nu \neq 0$, i.e. when friction is present; in this paper κ_r is called the "index of attenuation." κ_r

The basic modes of propagation fall into three types, which are distinguished by the nature of M_r^2 in the following way. M_r

(i) When M_r^2 is real and positive, μ_r has a non-zero value and κ_r is zero; this corresponds to the propagation of progressive waves along the positive or negative direction of the z-axis with phase velocity c/μ_r and constant amplitude. (ii) When M_r^2 is complex, κ_r is no longer zero and the progressive waves propagated, as in (i), in the positive and negative directions of Oz suffer exponential diminution of amplitude. This falls to $1/e$ of its value in a distance κ_r^{-1}. (iii) When M_r^2 is real and negative, μ_r is zero but κ_r has a non-zero value and the vibrations of a scalar quantity y take the form y

$$y = \cosh \kappa_r z \cos pt \mp \sinh \kappa_r z \sin pt$$

as compared with the form

$$y = \cos \kappa_r z \cos pt \mp \sin \kappa_r z \sin pt$$

of the progressive waves of (i). This is the type of disturbance which penetrates a medium where $M_r^2 < 0$. Such a medium is usually taken as totally reflecting, but, if it is finite in thickness, some penetration through it can take place, especially if its thickness is small compared with a vacuum wave-length. The limiting condition for $M_r^2 < 0$ is given by the point on the dispersion curve at which μ_r vanishes.

We note that, since actual absorption of energy can only take place when $\mu_r \kappa_r \neq 0$*, the second type of vibration is the only one in which energy is absorbed. $\mu_r \kappa_r$ is, save for the factor $2c/p$ depending on the frequency, the imaginary part of the right-hand side of (1), and cannot therefore have a non-zero value unless friction is present.

From the above discussion it is clear that, in order to predict the nature of the wave which will be propagated under any given conditions, and to discover what will happen to any wave incident on the ionized medium, we must first investigate the properties of the Appleton-Hartree formula and the types of dispersion curve given thereby. It may be well to emphasize here that the formula does not refer to media in which N is varying, but only to a uniform medium with constant N; it gives the types of vibration propagated in such a medium extending to infinity in the plane of the wave-front.

Discussion of a medium with variable N is only possible by working with the appropriate differential equations assuming the N-distribution corresponding to the medium; analytical solution of the equations in terms of known functions has not yet been performed except with a stepped N-distribution, consisting of a system of regions of finite extent in each of which N is constant. In this case the

* Cf. Drude, *Theory of Optics*, pp. 366–367.

values of M_r given by the Appleton-Hartree formula may be used to build up a wave in each separate region as described above. The unknown constants in this wave are determined by satisfying the boundary conditions. The reflected wave corresponding to an incident wave of given frequency, amplitude and polarization may be determined as regards amplitude, phase and state of polarization.

§ 3. TYPES OF DISPERSION CURVE

The physical quantities which define the medium under discussion are the longitudinal and transverse components H_L and H_T of the imposed magnetic field, the collision-frequency ν, the angular frequency p of the waves under investigation, and the density N of the charged scattering centres, which we will henceforth assume to be electrons. With so many variables there are several possible types of dispersion curve or surface and, in deciding which shall be most fully discussed, we must consider which are likely to be of the greatest utility to those engaged in experimental researches on the propagation of waves. The chief application of the Appleton-Hartree formula is to the interpretation of the results of experiments on the electrical constitution of the ionosphere. The formula was deduced by considering the case of waves incident normally on the ionized medium, and many experiments have been and are being done by emitting waves very close to the receiver and observing the wave reflected vertically downwards by the ionosphere. For such emissions H_L and H_T are fixed by the elements of the earth's magnetic field in the region where the experiments are carried out, since the emitter and receiver are very nearly at the same place. The collision frequency ν may next be fixed; it will be zero in the absence of friction, and a set of curves with values of ν increasing from zero will show the gradually increasing effect of friction on the properties of the medium.

The only independent variables remaining when the value of ν has been fixed are the angular frequency p and the electron density N. Experiments are generally designed to send a wave of given frequency into the ionosphere, where it meets a dispersive medium of gradually increasing N; thus, if we wish to be able to interpret the constitution of the reflected wave, we must draw the dispersion curves corresponding to the experimental arrangement—that is, we must fix p and find how the two values M_r^2 vary as N gradually increases from zero. The curves will then represent the conditions actually met by the wave as it penetrates into the medium, and also those encountered by the reflected wave on its return to the receiver. A set of such curves will be sections of three-dimensional (p, N, M_r^2) surfaces; sections of the surfaces at right angles to those from which they were built up will be dispersion curves giving the variation of M_r^2 with p for constant N. The dispersion curves given in this part of the paper are restricted to those for which $\nu = 0$; curves giving the values of μ_r and κ_r for three degrees of damping of different orders of magnitude, for four typical wave-lengths, with the corresponding polarizations, will form the subject of part 2.

§ 4. CASE OF ZERO DAMPING: CLASSIFICATION OF (N, M_r^2) DISPERSION CURVES

In the case where the electrons are not subject to collisional friction, we set $\nu = 0$ and therefore $\beta = 0$ in the expression on the right-hand side of (1). Both

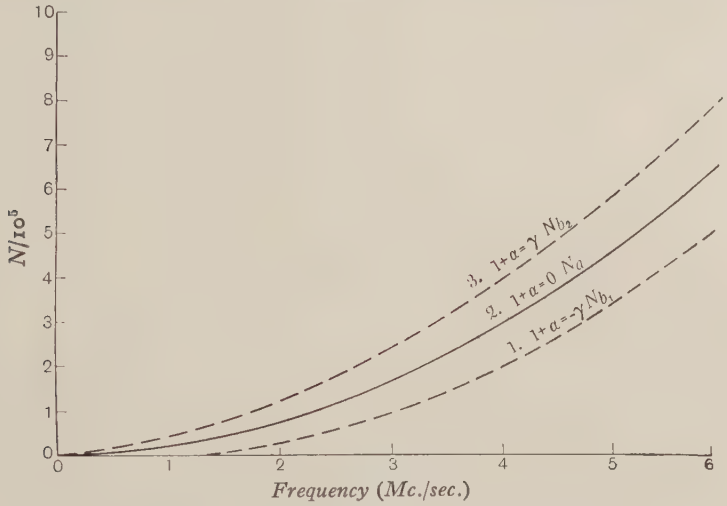


Figure 1. (Frequency, N) curves for $M_r = 0$.

values of M_r^2 now remain real, though they may be positive, negative or zero, and the only modes of propagation which can occur are (i) and (iii), described in § 2.

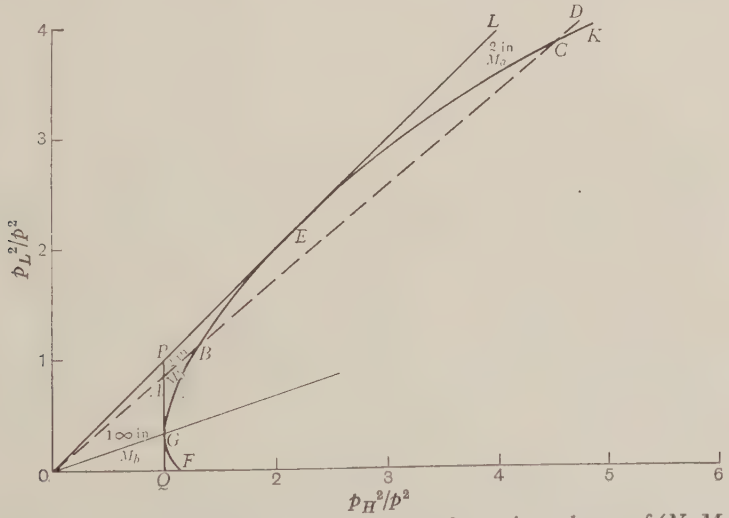


Figure 2. Ranges of magnetic field and frequency for various classes of (N, M_r^2) curves.

These modes are separated by the points, shown in figure 1, at which $M_r^2 = 0$. Discussion of the use of the curves giving $M_r^2 = 0$ is deferred until § 6, as their interpretation requires the use of the actual dispersion curves, which will now be described.

Since M_r^2 is now always real, it is convenient to represent μ_r and κ_r on one diagram, by drawing the curves for M_r^2 as functions of N, and remembering that points above the line $M_r^2 = 0$ give μ_r^2 and refer to mode (i), while points below this line give $\kappa_r^2 c^2/p^2$ and refer to mode (iii).

The curves showing M_r^2 as a function of N for constant p may be divided into four classes according to the types of infinity they contain. Both values of M_r^2 never become infinite for one value of p; the infinities, when they exist, lie either in M_a^2 or in M_b^2. The values of N for which it is possible for a value of M_r^2 to become infinite are given by the zeros of the denominator of the Appleton-Hartree formula, which correspond to positive values of N, that is, by the real positive roots of the cubic equation in p_0^2/p^2:

$$\left(\frac{p_0^2}{p^2}\right)^3 + \frac{9}{2}\left(\frac{p_0^2}{p^2}\right)^2 + \frac{27}{2}\left(\frac{p_0^2}{p^2}\right)\left(\frac{p_H^2}{3p^2} - \frac{p_L^2}{p^2}\right) + \frac{27}{2}\left(\frac{p_H^2}{p^2} - 1\right) = 0 \quad \ldots\ldots(3).$$

Taking Cartesian co-ordinates x, y where $x = p_H^2/p^2$, $y = p_L^2/p^2$, we can represent in one diagram, figure 2, the ranges of magnetic field and frequency for which the various classes of (N, M_r^2) dispersion curves occur. Since p_H^2/p^2 and p_L^2/p^2 both > 0 and $p_L \leqslant p_H$, we are restricted to the first octant of the $(p_H^2/p^2, p_L^2/p^2)$ plane. Any line through the origin defines a value of p_L/p_H, that is, of the angle of dip, so that when considering the curves applying to any given place we move along one particular straight line through O; the angular frequency p is infinite at the origin and decreases as we move outwards. At Slough (latitude 51½° N., longitude 0°30′ W.) we are on the line OD; it is with reference to this line that the curves shown have been drawn and, as it is quite typical of places for which $p_L^2 > \frac{1}{3}p_H^2$, the general phenomena will for simplicity be described with reference to it. From the absolute value of the strength of the earth's magnetic field, we can find the point on OD corresponding to any particular frequency and, conversely, the frequency corresponding to any point on OD which we may find particularly interesting.

Consideration of the properties of the roots of (3) shows that there is one infinity in M_b^2 for all points for which $p_H^2/p^2 > 1$. The line $p_H^2/p^2 = 1$ corresponds to the natural frequency of rotation of the electrons around the total magnetic field and gives at A on OD a wave-length 227·1 metres. Class 1 of dispersion curve (see figure 3·1, 1–5) is obtained for all wave-lengths shorter than this; the infinity in M_b^2 moves towards larger values of N as the frequency increases and dM_b^2/dN always < 0. At A another infinity in M_b^2 appears at $N = 0$ and, as the frequency decreases, moves towards regions of greater N until at B the two infinities in M_b^2 coincide.

The curve for 227·1 m., figure 3·1, 6, shows the boundary curve between classes 1 and 2. The mathematical infinity appearing here at the origin in M_b^2 has no real physical meaning; the theory giving the Appleton-Hartree formula is based on the assumption that a large number of electrons is contained in a cube of the medium whose side is one vacuum wave-length, so that it does not strictly apply in the immediate neighbourhood of $N = 0$. The dispersion curves must all start from $M_r^2 = 1$ for $N = 0$; exact investigation of physical behaviour in the neighbourhood

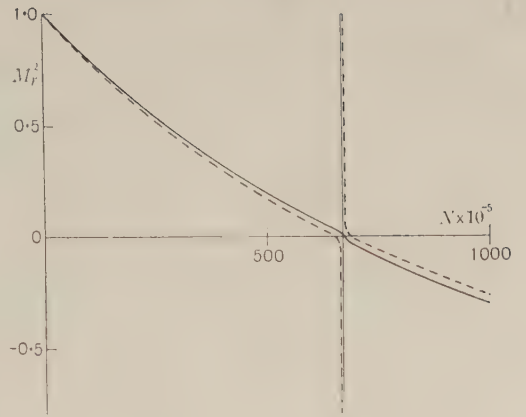


Figure 3·1, 1. $\lambda = 5$ metres.

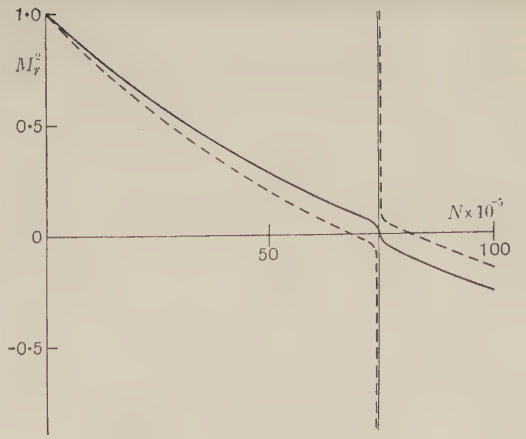


Figure 3·1, 2. $\lambda = 15$ metres.

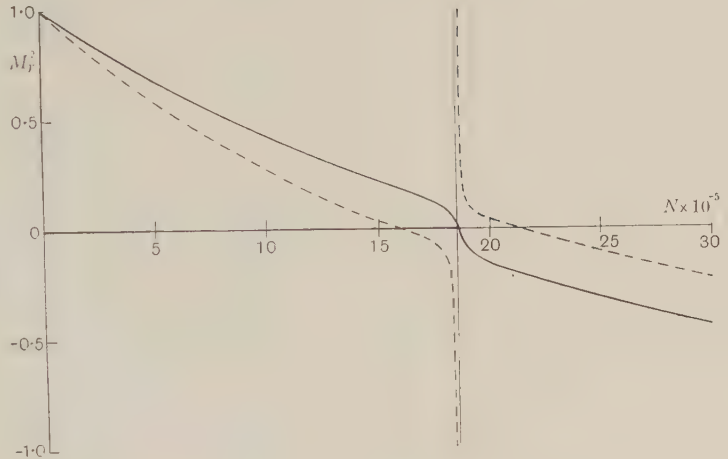


Figure 3·1, 3. $\lambda = 30$ metres.

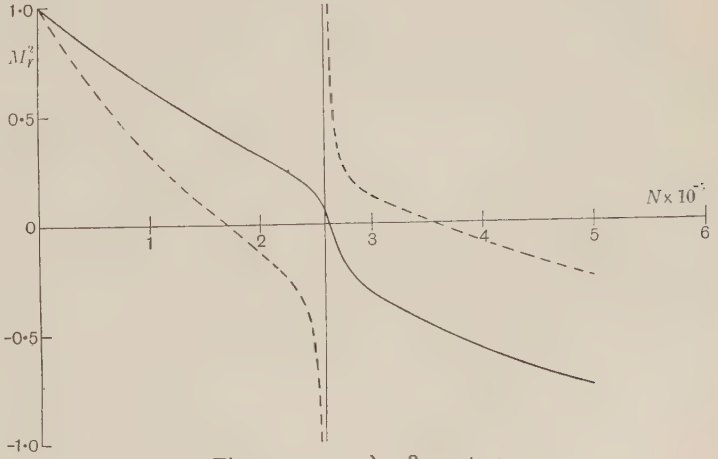


Figure 3·1, 4. $\lambda = 80$ metres.

Figures 3·1, 1 to 3·1, 4. (N, M_r^2) curves. M_a^2 ——— M_b^2 ------

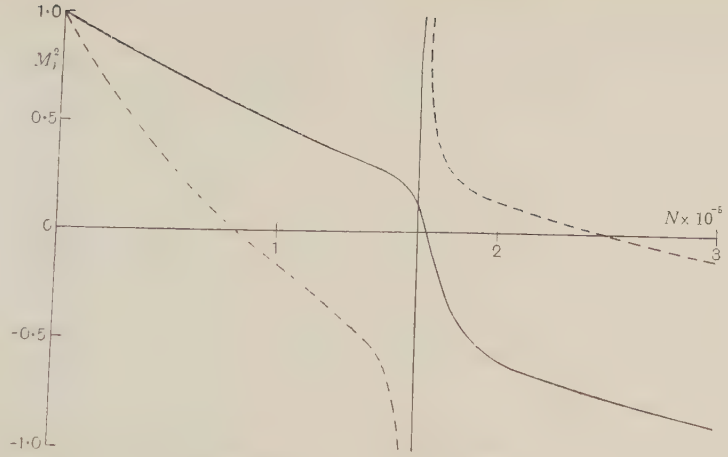


Figure 3·1, 5. $\lambda = 100$ metres.

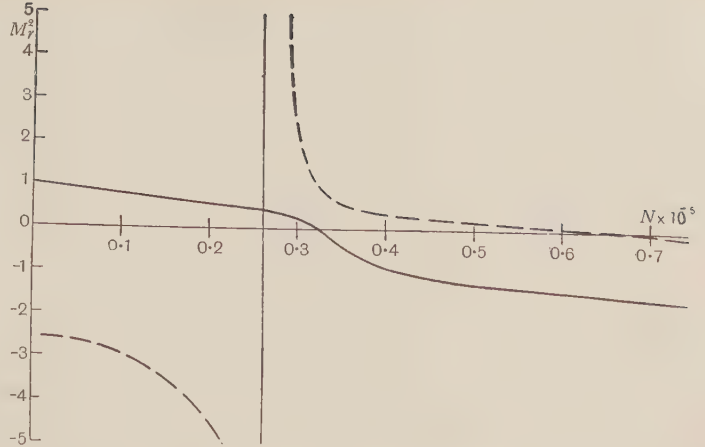


Figure 3·1, 6. $\lambda = 227{\cdot}1$ metres.

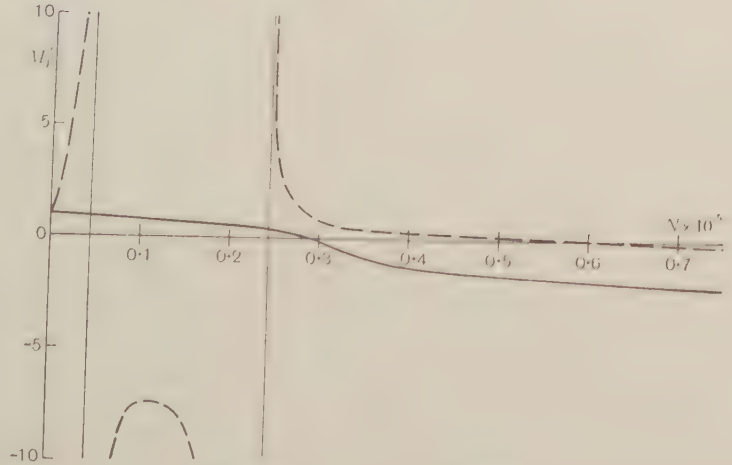


Figure 3·1, 7. $\lambda = 240$ metres.

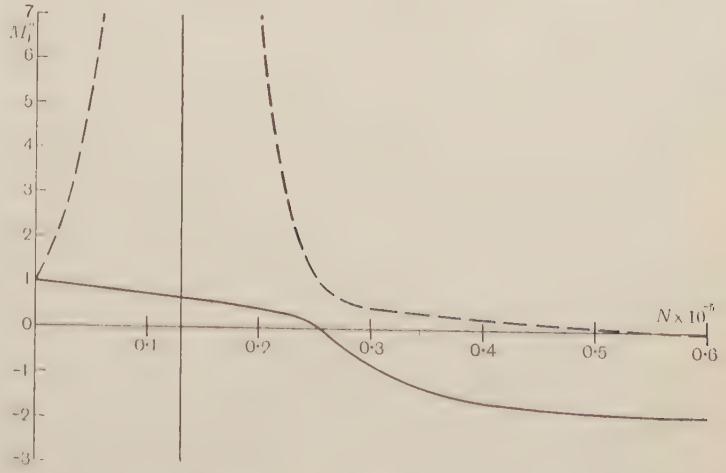


Figure 3·1, 8. $\lambda = 256{\cdot}2$ metres.

Figures 3·1, 5 to 3·1, 8. (N, M_r^2) curves. M_a^2 ——— M_b^2 - - - - - -

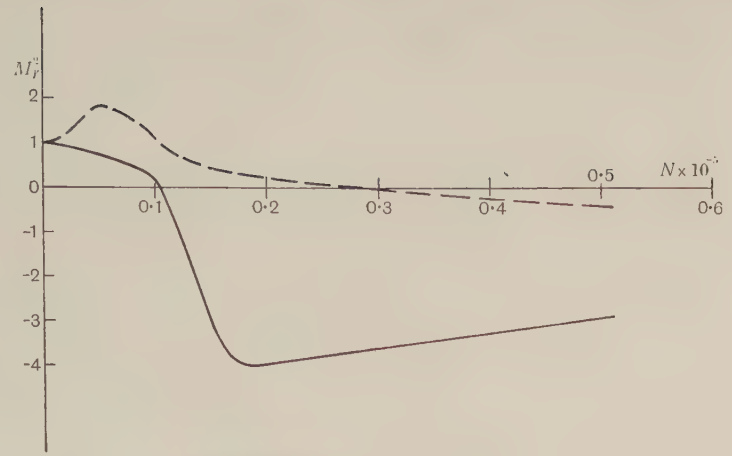


Figure 3·1, 9. $\lambda = 400$ metres.

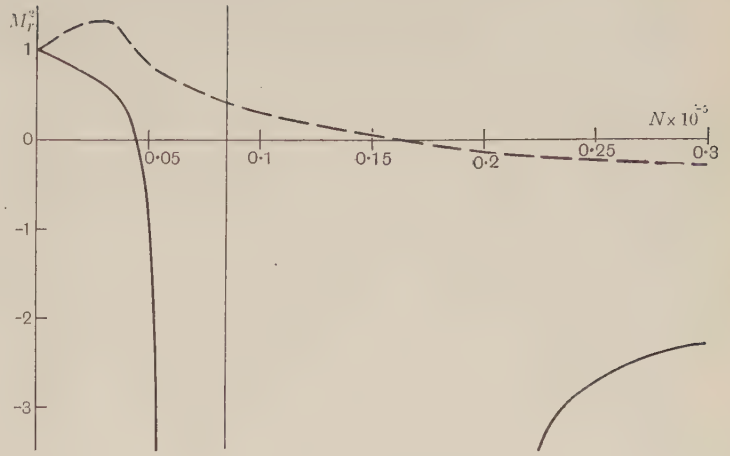


Figure 3·1, 10. $\lambda = 615{\cdot}5$ metres.

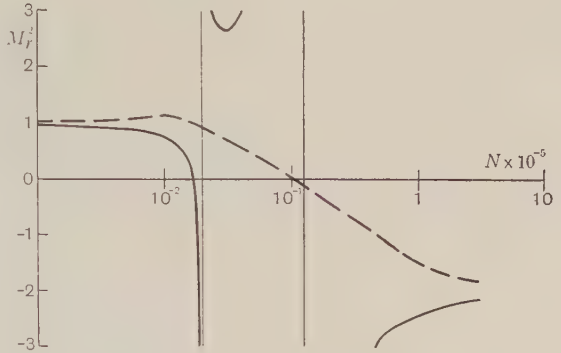


Figure 3·1, 11. $\lambda = 1000$ metres.

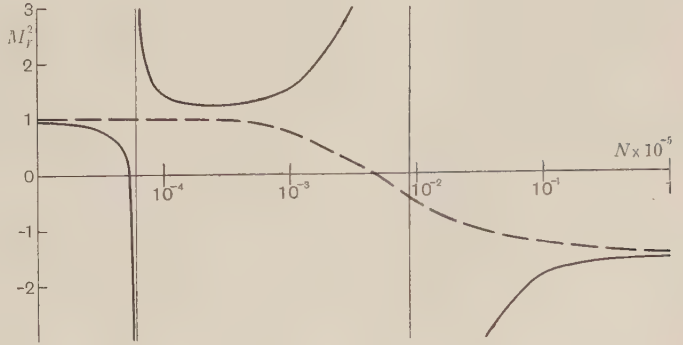


Figure 3·1, 12. $\lambda = 18{,}000$ metres.

Figures 3·1, 9 to 3·1, 12. (N, M_r^2) curves. M_a^2 ——— M_b^2 ------

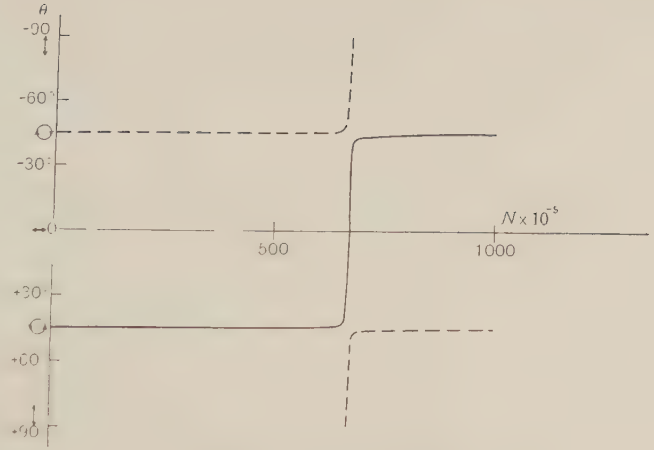


Figure 3·2, 1. $\lambda = 5$ metres.

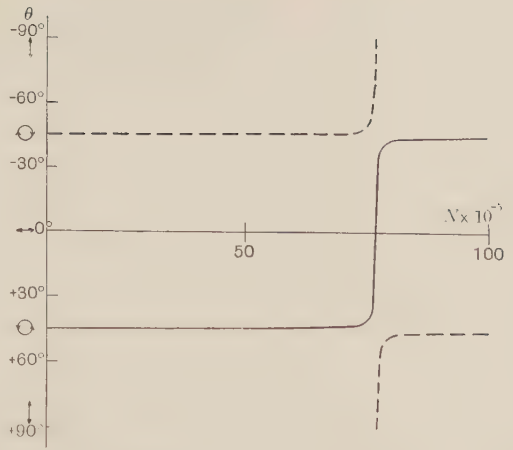


Figure 3·2, 2. $\lambda = 15$ metres.

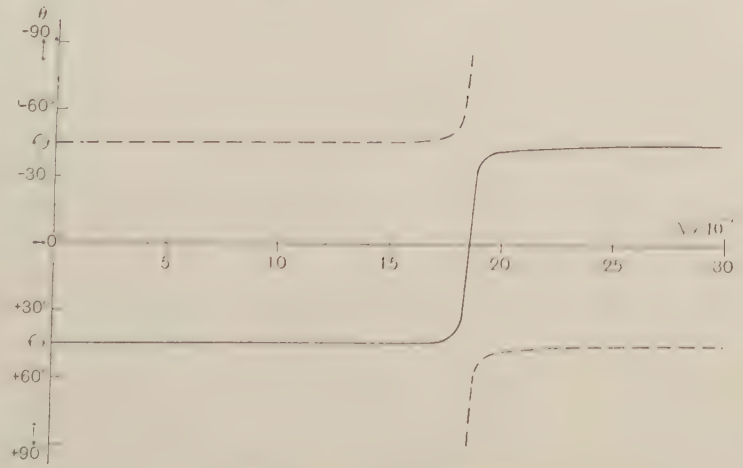


Figure 3·2, 3. $\lambda = 30$ metres.

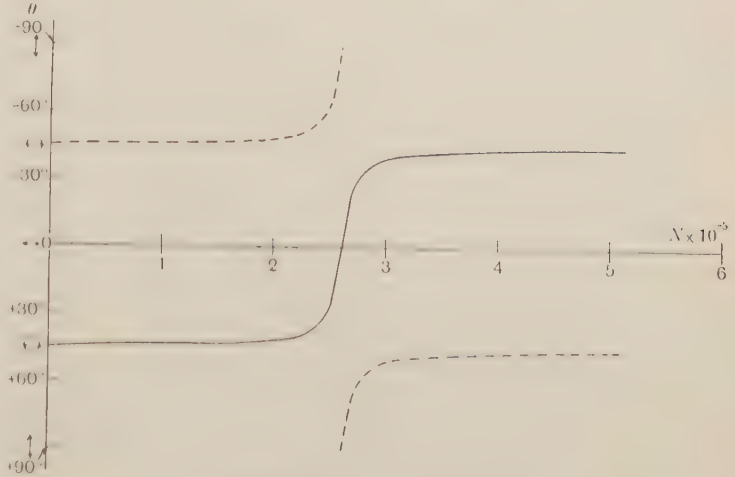


Figure 3·2, 4. $\lambda = 80$ metres.

Figures 3·2, 1 to 3·2, 4. Polarization curves. M_a^2 ——— M_b^2 - - - - - -

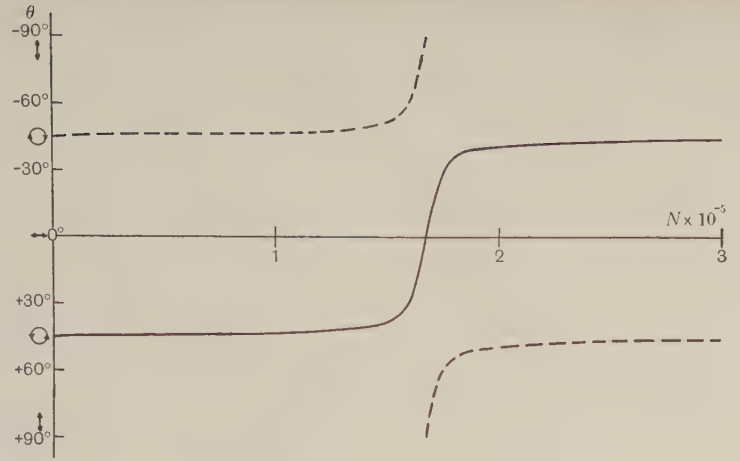


Figure 3·2, 5. $\lambda = 100$ metres.

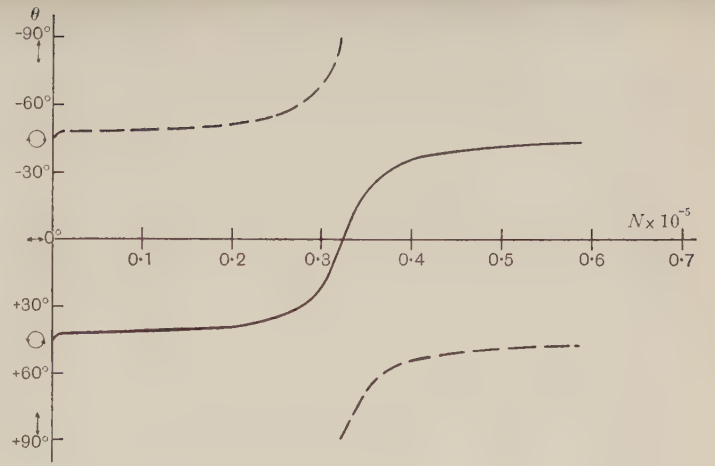


Figure 3·2, 6. $\lambda = 227{\cdot}1$ metres.

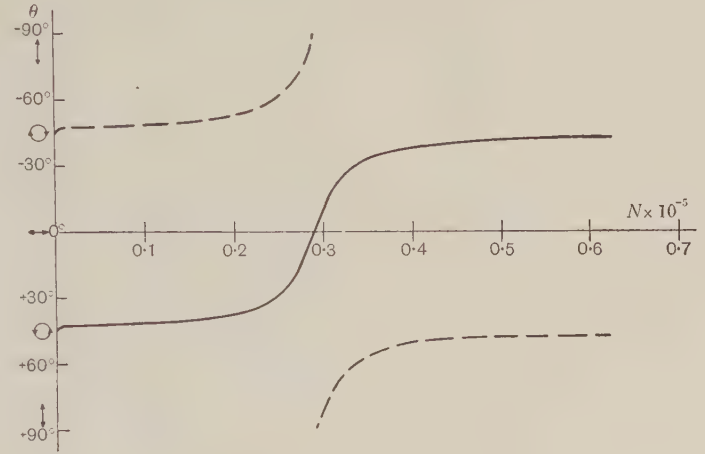


Figure 3·2, 7. $\lambda = 240$ metres.

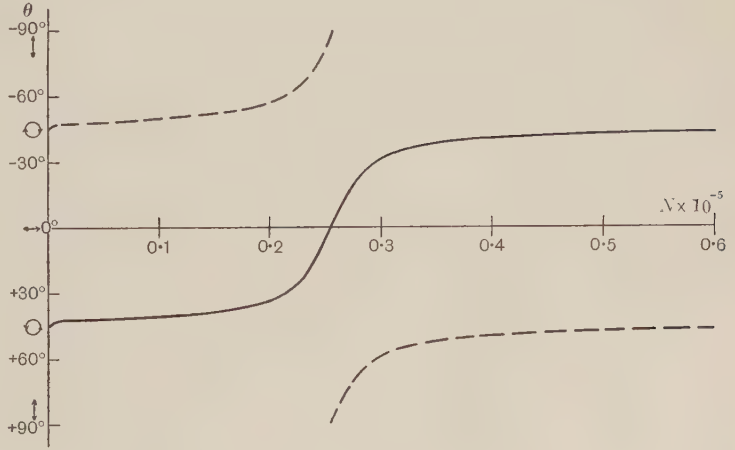


Figure 3·2, 8. $\lambda = 256{\cdot}2$ metres.

Figures 3·2, 5 to 3·2, 8. Polarization curves. M_a^2 ——— M_b^2 - - - - - -

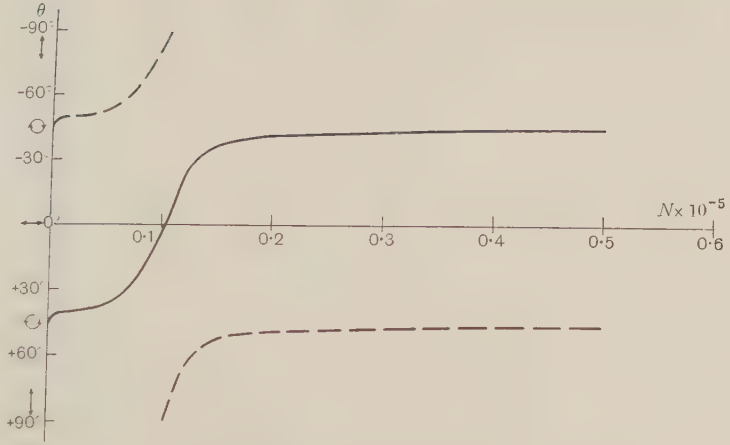


Figure 3·2, 9. $\lambda = 400$ metres.

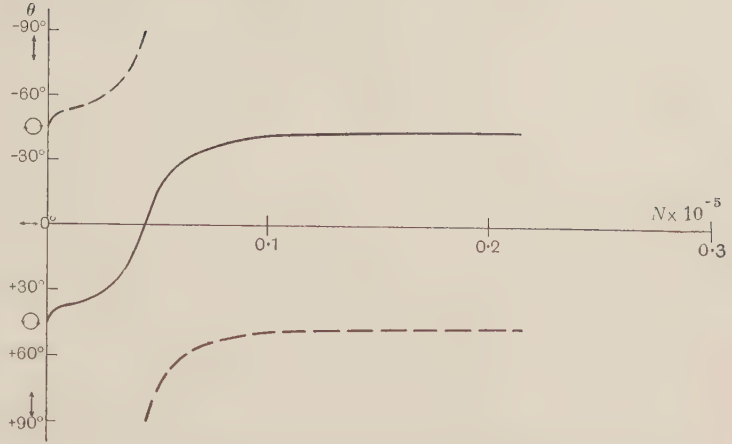


Figure 3·2, 10. $\lambda = 615{\cdot}5$ metres.

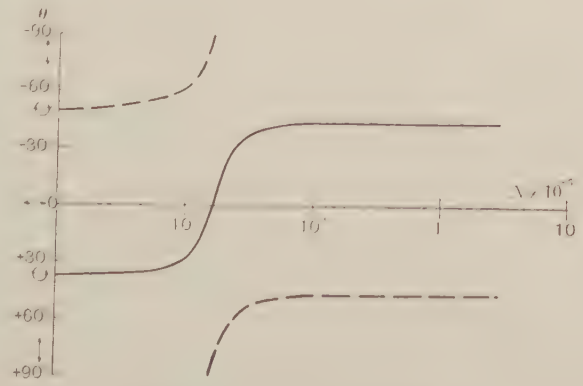


Figure 3·2, 11. $\lambda = 1000$ metres.

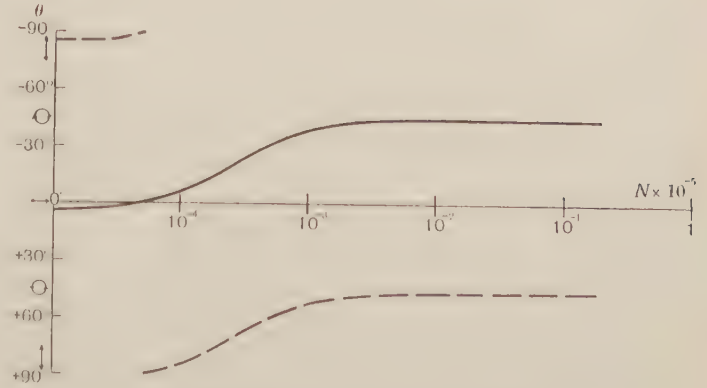


Figure 3·2, 12. $\lambda = 18{,}000$ metres.

Figures 3·2, 9 to 3·2, 12. Polarization curves. M_a^2 ———— M_b^2 ------

of $N = 0$ would require examination of the validity near this point of the fundamental assumptions of the theory, and falls outside the scope of this paper. The M_b^2 curve for $\lambda = 227{\cdot}1$ m. has been drawn starting at a finite distance from $N = 0$.

This difficulty that the curve for $p = p_H$ gives $M_b^2 \to -\infty$ in the limit for $N \to 0$ is not peculiar to the problem of propagation of electromagnetic waves in an ionized medium in the presence of an external magnetic field but occurs at resonance in any dispersion problem, e.g. in the case of classical oscillators. In the Lorentz form of dispersion theory, the value $M^2 = \infty$ at resonance occurs however few the oscillators, and one gets $M^2 \to \infty$ as $N \to 0$, though, putting first $N = 0$ and then $p \to$ resonance frequency, one gets $M^2 = 1$. Care is thus necessary in inversion of limits, since in our problem we have

$$\lim_{N\to 0} \lim_{p\to p_H} M_b^2 \neq \lim_{p\to p_H} \lim_{N\to 0} M_b^2.$$

In the present case it is the second-order terms involving the magnetic field which prevent M_b^2 having an infinite value for all N at this frequency.

The introduction of even a small amount of friction causes the infinity to disappear, just as in the ordinary Lorentz dispersion formulae.

The curve for 240 m., figure 3·1, 7, is given as an illustration of class 2, with two infinities in M_b^2. The curve for 256·2 m., figure 3·1, 8, is the boundary between classes 2 and 3; $M_b^2 \to \pm\infty$ on both sides of the value of N for which it becomes infinite. Between B and C neither M_a^2 nor M_b^2 has any infinities, since equation (3) has no real positive roots. This class 3 is illustrated by the curve for $\lambda = 400$ m., figure 3·1, 9. At C, corresponding to $\lambda = 615{\cdot}5$ m., figure 3·1, 10, for which the curve is shown, a double infinity appears in M_a^2 and for lower frequencies this separates into two, one of which moves towards very small values of N. This class 4 is illustrated by the curves for $\lambda = 1000$ m. and $\lambda = 18{,}000$ m., figure 3·1, 11 and 12.

The curve in the (x, y) plane which separates the regions of two and zero infinities is the cubic curve *FGBECK*, whose equation is

$$8\,(\tfrac{1}{3}x - y)^3 - \tfrac{1}{3}x^2 + 14xy - 3y^2 - 12y = 0 \qquad \text{.....(4).}$$

It is tangential to $x = 1$ at G, where $y = \frac{1}{3}$, and to the line $y = x$ at E, where $x = y = \frac{11}{14}$. In the whole of the region *OPQ* there is one infinity in M_b^2; in *GPE* there are two infinities in M_b^2, and in *LEK* two in M_a^2. Within the whole of the region lying between *QGBECK* and the line $x = 0$ there are no infinities in either M_a^2 or M_b^2. Thus at places, such as those near the magnetic equator, for which $p_L^2 < \frac{1}{3}p_H^2$, there are only two classes of (N, M_r^2) dispersion curves, one with one infinity in M_a^2 and the other with no infinities in either M_a^2 or M_b^2. These are the Slough classes 1 and 3; see, for instance, figure 3·1, 4 and 9. At places where p_L^2 lies between $\frac{1}{3}p_H^2$ and p_H^2, there are the four classes of curve described above with reference to the line *OD*. Thus figure 2 enables us at once to determine the type of dispersion curve holding for any given frequency at any given place.

For the line OD the point B corresponds to $\lambda = 256{\cdot}2$ m. and C to $\lambda = 615{\cdot}5$ m. That these values coincide very nearly with p_L and p_T, the natural frequencies of rotation about the longitudinal and transverse components of the magnetic field respectively, is seen from figure 2 to be due to the value of the angle of dip at Slough rather than to any deeper physical cause. The cubic curve represented by equation (4) seems to have no simple physical significance.

§ 5. NOTES ON THE DISPERSION CURVES ILLUSTRATED

(N, M_r^2) dispersion curves for vertical reflection from the ionosphere at Slough are given in figure 3·1 for the wave-lengths shown in the figure. They fall into the following classes: Class 1, one infinity in M_b^2, figures 3·1, 1 to 3·1, 5. Boundary curve, figure 3·1, 6. Class 2, two infinities in M_b^2, figure 3·1, 7. Boundary curve, figure 3·1, 8. Class 3, no infinities, figure 3·1, 9. Boundary curve, figure 3·1, 10. Class 4, two infinities in M_a^2, figures 3·1, 11 and 3·1, 12.

Apart from the boundary curves between the classes, the wave-lengths illustrated have been chosen with a view to demonstrating the effect of the presence of a magnetic field on propagation at the various ranges of frequency which are roughly divided off by the experience of practical wireless transmission. It is hoped that the curves may prove to be of practical value to those engaged in experiments on the nature of radio propagation. Each curve has been treated on its own merits and a scale has been chosen for it which will bring out its peculiar properties (e.g. the points where $M_r^2 = 0$, the various branches and asymptotes) as clearly as possible. The N and M_r^2 scales are therefore not uniform throughout the set; the N scale in particular becomes very much more spread out as the frequency decreases. For the long-wave-length curves of class 4 it was necessary to choose a logarithmic scale in order to show the behaviour for small values of N.

The effect of the magnetic field does not begin to be noticeable for the values of N, greater than 10^6, which are believed to exist in the ionosphere, until the wave-length *in vacuo* has increased from zero to the order of 30 m., figure 3·1, 3. For longer wave-lengths the effect becomes stronger, and for $\lambda = 80$ and 100 m., figures 3·1, 4 and 5, the dispersion is very marked from $N = 0{\cdot}5 \times 10^5$ to $N = 4 \times 10^5$, so that the magnetic field may be expected to have a large influence on propagation in this region. As λ increases, the range of N in which the effect of the magnetic field is important moves towards smaller values.

§ 6. USE OF THE ZEROS AND INFINITIES OF THE DISPERSION CURVES IN THE INTERPRETATION OF PROPAGATION PHENOMENA

Appleton and Hartree have shown that the points where $M_r^2 = 0$ are determined by

$$\left.\begin{aligned} 1 + \alpha &= 0 \text{ for } M_a^2 \\ 1 + \alpha &= \pm\gamma \text{ for } M_b^2 \end{aligned}\right\} \quad \ldots\ldots(5),$$

i.e.

$$\left.\begin{aligned} p_0^2 &= \frac{3}{2}p^2 \quad \text{for } M_a^2 \\ \frac{2}{3} - \frac{p^2}{p_0^2} &= \pm\frac{pp_H}{p_0^2} \text{ for } M_b^2 \end{aligned}\right\} \qquad \ldots\ldots(5a).$$

We notice that $M_b^2 = 1$ for $M_a^2 = 0$, when $N = N_a$, say.

Since $p_0^2 = 4\pi Ne^2/m$, these equations give the critical values of N for which $M_r^2 = 0$ at any given frequency. They are shown in figure 1 as functions of the frequency.

Curve 2, figure 1, shows N_a, the value corresponding to $M_a^2 = 0$, which is independent of the magnetic field and exists for all frequencies; N_b, corresponding to $M_b^2 = 0$, has two values: N_{b_1} shown in curve 1, figure 1, corresponding to the equation $1 + \alpha = -\gamma$, and N_{b_2}, shown in curve 3, figure 1, corresponding to the equation $1 + \alpha = +\gamma$. For $p < p_H$ there are no values of N_{b_1}. In considering which of the curves in figure 1 is likely to give a point of reflection, we must also consult the actual dispersion curves given in figure 3·1.

From these we see that, for all angular frequencies greater than p_H, i.e. for all wave-lengths less than 227·1 m. (determined by the value of the earth's magnetic field prevailing at Slough) the wave corresponding to M_a^2 (frequently called the "ordinary wave") may be expected to be reflected by values of N given by N_a, while that corresponding to M_b^2 (frequently called "extraordinary wave") may be expected to be reflected by N_{b_1} as it goes up to meet increasing values of N. The disturbance beyond N_{b_1} produced by the wave reflected there will in general have diminished so much in amplitude before it meets densities at which its propagation is again of type (i), § 2 (between the infinity line and N_{b_2}) that no appreciable amount of energy remains to be reflected by N_{b_2}. We may therefore expect the extraordinary wave to be reflected before the ordinary wave at these frequencies, and so to have a shorter delay-time when the reflection of pulses is being studied. If, however, the variation of N is so rapid that the disturbance is still appreciable when N attains the value N_{b_2}, it would be reflected there but would probably not have sufficient amplitude after reflection and attenuation to cause an appreciable disturbance in the region $N < N_a$ on the return journey. There is, however, the theoretical possibility that it might do so and be reflected as an echo of very long delay.

For longer wave-lengths the phenomena become more complicated. At frequencies belonging to class 2, for instance $\lambda = 240$ m., figure 3·1, 7, there is a possibility of reflection of the extraordinary wave before either of the zeros N_a or N_{b_2} is reached, owing to the early occurrence of the first infinity in M_b^2. The ordinary wave will still be reflected from its zero at N_a, but a rapid increase in N will be required to enable the extraordinary wave disturbance to pass through to N_{b_2} and be reflected there. At the boundary between classes 2 and 3, where $\lambda = 256{\cdot}2$ m., the total reflection for both the ordinary and extraordinary waves occurs at the zeros of M_r^2, but now the curves to use in figure 1 are 2 and 3 instead of 1 and 2

as they were in class 1; we should now expect the ordinary wave to give the reflected pulses of lesser delay. We may here notice that $M_b^2 > 1$ for values of N up to N_a in curves of type 3 means simply that the phase velocity of propagation of the extraordinary wave is less than the velocity of light for such values of N; this holds for any N where $M_b^2 > 1$. This behaviour is typical also of class 3 and the boundary curve between classes 3 and 4. With class 4 there is in addition the possibility that the ordinary wave disturbance may pass through the very small variation in N required to make its propagation possible again. On these long wave-lengths it seems quite probable that transmission of the ordinary wave may take place through its first totally reflecting region; it would then be reflected from its second infinity point as a long-delay echo. The behaviour of M_b^2 is regular in this case; it is simply totally reflected from N_{b_2}.

The values of N which are required to reflect waves of class 4 are, however, so small that the absorption must be very large in the region where they are reflected; this will probably have the effect of smoothing out the infinities and bumps on the curves, so that the small regions of possible propagation near the asymptotes are practically eliminated. For these long wave-lengths there may be a true discontinuity and an effective sharp boundary to the reflecting layer, since continuity of N physically means a small change in a wave-length.

§ 7. (p, N) REGIONS OF PROGRESSIVE PROPAGATION OF THE BASIC MODES

The complicated reflection phenomena described in the preceding paragraph are summarized in figure 4, which shows the division of the (p, N) plane into

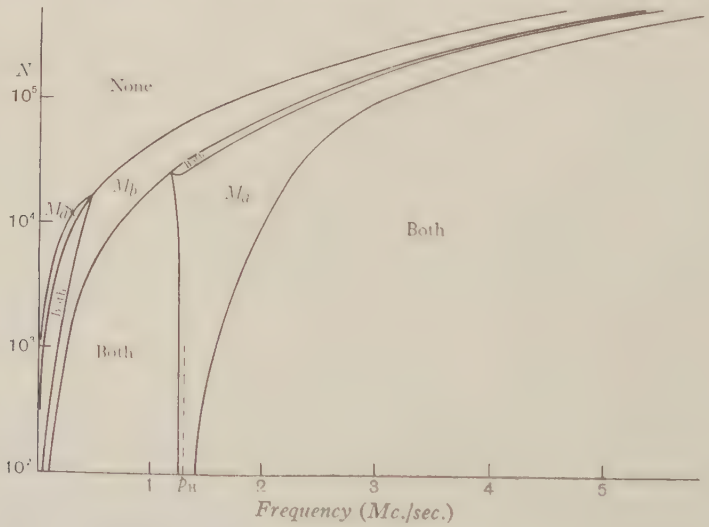


Figure 4. Division of (p, N) plane into regions of progressive propagation of the basic modes.

regions of progressive propagation of the basic modes, i.e. regions which give type (i), § 2, for one or other of the modes. A logarithmic N scale spreads out the

curves for small p and N which are all asymptotic to the N-axis; in drawing the curves in this region, it was thought important to stress their relative positions rather than to adhere closely to the calculated values. A line drawn parallel to the N-axis for any given frequency will show the successive propagation conditions encountered by upgoing waves of that frequency; in using the diagram it should be remembered that, with a sufficiently large variation of N within a wave-length, it may be possible for the wave disturbance to pass through the totally reflecting region into a region where it can be propagated again, and to suffer reflection at the upper boundary and perhaps further reflection back and forth until it emerges after a long delay or is too attenuated to be detected.

§ 8. POLARIZATION

Equation (2) determines the polarizations corresponding to the two values of M_r^2. Those corresponding to the (N, M_r^2) curves of figure 3·1 are shown graphically in figures 3·2, 1 to 3·2, 12. The angle θ which is shown as a function of the electron-density is defined by the equation

$$\tan\theta = iEy/Ex = Hx/iHy,$$

where the axes Ox and Oy are respectively along and perpendicular to the transverse component of the earth's magnetic field; $\tan\theta$ thus gives the ratio of the axes of the ellipse described by the electric vector in the (x, y) plane. Linear polarization along Ox is given by $\theta = 0$ and along Oy by $\theta = \pm 90°$; positive values of θ correspond to an anticlockwise (left-hand) direction of rotation viewed along the direction of propagation, and negative values of θ to clockwise (right-hand) rotation*. $\theta = 45°$ corresponds to left-hand circular and $\theta = -45°$ to right-hand circular polarization. For values of θ between 0° and $\pm 45°$ the ellipse has its major axis along Ox, for values of θ between $\pm 45°$ and $\pm 90°$ along Oy. We note that there is a possible ambiguity of θ with respect to whole multiples of 180°.

In every case the limiting polarizations corresponding to M_a^2 and M_b^2 as $N \to 0$ are circular and are left-handed and right-handed respectively, when the positive direction of H_z is along the positive direction of propagation of the mode considered†. These are the conditions which apply to the down-coming waves in the northern magnetic hemisphere. The diagrams show that as N increases the polarizations remain very nearly circular until the wave reaches the immediate neighbourhood of the point N_a where $M_a^2 = 0$, $M_b^2 = 1$. Then they rapidly become elliptical; the y-axes of the ellipses corresponding to M_a^2 and M_b^2 become relatively smaller and greater respectively, until at the point $N = N_a$ the polarizations are linear, along the x- and y-axes respectively. For greater values of N they open out and soon become nearly circular again. The polarization corresponding to M_a at the points N_{b_1}, N_{b_2}, where $M_b^2 = 0$, is given by $Hy/Hx = \pm i\gamma_L/\gamma$ respectively, so that θ is a function of the angle of dip alone.

* Note that in figures 3·2, 1–12 the positive values of θ are below the N-axis.

† The charge on an electron, the sign of which determines that of γ_L, is negative.

In the present case of zero absorption, the principal axes of the ellipse remain along the x- and y-axes. The axes are the same for the two waves and the direction of description is opposite in the two cases. Their ellipticities are the same, but the major axis of the one is the minor axis of the other. Mathematically speaking, the product of the polarizations at any point is -1 and the two values of θ are complementary in magnitude. The main effect of friction on polarization is to rotate the principal axes through an angle which may be calculated for any point; the product of the polarizations is still, however, -1. This subject will be discussed in Part II.

§ 9. ACKNOWLEDGMENTS

This work was carried out at the Radio Research Station, Slough, as part of the programme of the Radio Research Board of the Department of Scientific and Industrial Research, and is published by permission of the Board. Thanks are due to Mr R. A. Watson Watt, Superintendent of the Station, for his interest in the work and for the provision of facilities for its conduct; to Prof. E. V. Appleton and Prof. D. R. Hartree for valuable discussion, and suggestions as to the presentation of the paper; to Miss A. C. Stickland for assistance in the calculations, and to Mr E. C. Slow for help in drawing out the curves.

REFERENCES

(1) E. V. Appleton. *Proc. Phys. Soc.* **37**, 16 D (1925).
(2) E. V. Appleton. *U. R. S. I. papers* (Washington 1927).
(3) D. R. Hartree. *Proc. Camb. Phil. Soc.* **27**, 143 (1931).
(4) S. Goldstein. *Proc. R. S.* A **121**, 260 (1928).

DISCUSSION

Mr R. A. Watson Watt. The author's interim report of progress in her heroic investigation of the theory of propagation of electromagnetic waves in an ionized medium under an external magnetic field is timely and valuable. I was, some years ago, so strongly impressed by the need for such a general investigation and by the formidable difficulties of bringing together adequate mathematical and physical skill and adequate computing facilities for bringing the problem to numerical solution that I suggested to the Radio Research Board that the work should be taken up at Slough, where close contact with experimental work would be a valuable guide to the mathematical investigator.

The results now presented, despite the limitations arising from neglect of collisional damping, are of a kind required by all investigators concerned with the mechanism of return of wireless waves from the ionosphere, and this paper will stand as a source from which initial data for new investigations can be obtained without duplication of effort.

It is satisfactory that we can report that the investigation of typical cases, taking account of collisional damping, is making satisfactory progress and may be reported at a comparatively early date.

Prof. E. V. Appleton. All experimental workers in the field of ionospheric investigations will welcome Dr Taylor's exhaustive representation of illustrative magneto-ionic dispersion curves. It is only by having the theory so fully worked out for us at the first stage that we are able to make the next step and consider the effects of collisional friction in differentiating between the attenuation experiences of the ordinary and extraordinary rays in their ionospheric journeys. I have always felt somewhat guilty concerning the complications which the gyroscopic terms introduce into the simple dispersion equations of Eccles and Larmor, but the experiments on the polarization of down-coming waves in England and Australia leave little doubt that we must face this difficulty.

Author's reply. The limitations arising from neglect of collisional damping in the curves given in the present paper are not apparent until the frequency of collisions is of the order of 10^5 per second, and even then the effect of friction is merely to round off the corners of the dispersion curves by removing the infinities and to prevent μ and κ from being ever actually zero. This degree of friction is most probably that pertaining to the Kennelly-Heaviside layer, at heights of 110–120 km. above the earth's surface, and the curves given in this paper may be confidently used in investigations of propagation phenomena at greater heights than this.

.539.132 531.12–2

A MODEL TO ILLUSTRATE THE MOTION OF A DIATOMIC ROTATOR WITH TWO DEGREES OF FREEDOM

By LEWIS SIMONS, D.Sc. and E. H. SMART, M.A.,
Birkbeck College, London

Received November 28, 1932. *Read, with demonstration, December* 16, 1932.

ABSTRACT. An arm 5 in. long is pivoted at one end and carries at the other end a small electric lamp which is thus capable of moving on the surface of a sphere about the pivot as centre. The two angular velocities $\dot{\phi}$ and $\dot{\theta}$ can be independently controlled by two hand regulated motors: ϕ is the azimuthal and θ the co-latitudinal angular co-ordinate of the arm. If $\dot{\phi}$ and $\dot{\theta}$ are commensurable, the resulting figure traced out by the lamp remains stationary in space. This path represents the motion of one of the atoms of the rotator which has two degrees of freedom.

The apparatus is designed to show a certain class of motions of a point which possesses two degrees of freedom. The type of motion is thus defined: The point moves in a circle while the circle rotates about a fixed diameter.

The point is thus constrained to move on the surface of a fixed sphere and the
a radius a to the point has two angular velocities $\dot{\theta}$ in the circle, where θ is measured from the fixed diameter, and $\dot{\phi}$ that of the plane of the circle about the fixed diameter.
θ We may conveniently take the fixed diameter as the polar axis: then θ is the co-
ϕ latitude and ϕ the azimuth measured from some fixed plane through the polar axis. It is convenient at the outset to take this fixed plane such that θ and ϕ vanish together.

n If $\dot{\theta} = n\dot{\phi}$ where n is some *commensurable* number the point P describes on the surface of a sphere a closed path which it continually retraces. This path may be
r readily obtained if we know its projection on the equatorial plane. If r is the radius vector of the projection of P in this plane, $r = a \sin \theta$. But, with the above-mentioned proviso, $\theta = n\phi$ if $\dot{\theta} = n\dot{\phi}$, and the equation of projection in the equatorial plane is (with ϕ as the vectorial angle) $r = a \sin n\phi$. This curve is characteristic of a class of plane polar curves known as rhodoneae from a fancied resemblance to rose petals. If n is an integer the curve consists of a number of loops passing through the origin, n in number if n is odd, $2n$ in number if n is even.

k If $\dot{\theta}$ and $\dot{\phi}$ are connected by a linear relation, say $\dot{\theta} = n\dot{\phi} + k$ where k is constant, the effect will be to rotate the path of the point about the polar axis with an angular
t velocity $(-k/n)$, for then $\theta = n\phi + kt$ (the constant of integration may be neglected).

The figures for a number of simple cases of commensurable values of n have been drawn in figures 2 (a) to 2 (f).

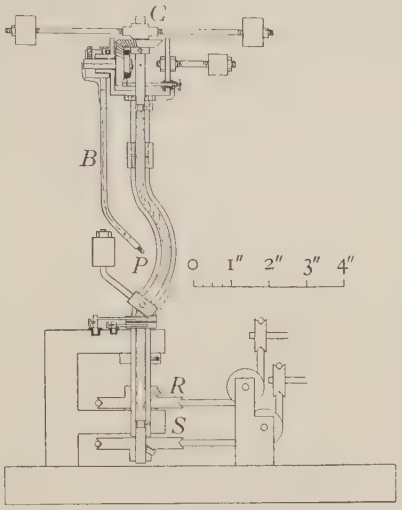


Figure 1.

The importance of this type of motion cannot be overstressed because of the attention given to it in work on molecular spectra. Here the motion occurs amongst their rotational degrees of freedom of variously constructed molecules, and, of

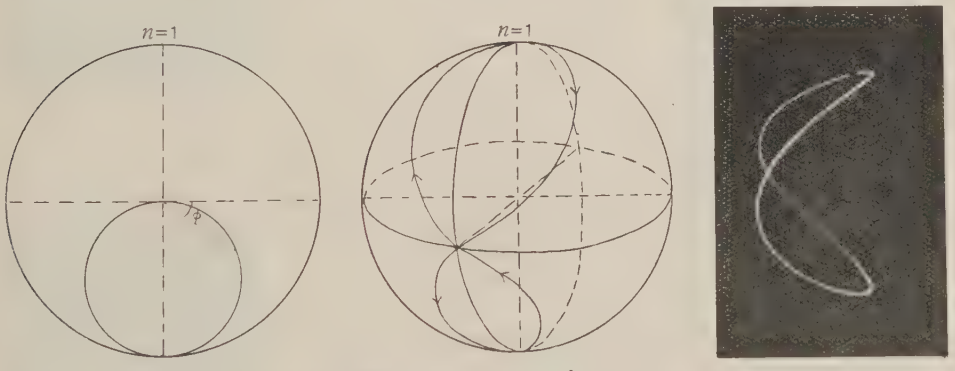


Figure 2 (a)*. $r = a \sin \phi$.

course, the picture has been made still more complex by quantizing the motion and by expressing it in terms of wave-mechanical theory†.

On account of the beauty of the motion, even with two degrees of rotational freedom, it was decided to try to construct a model to illustrate it, and figure 1

* The curve is the intersection of the sphere by the circular cylinder $x^2 + y^2 = ay$.

† The kinematics of a particle on the rim of a symmetrical top is a common enough subject, but the modifications introduced by quantum and wave-mechanical theory in direct application of the motion to atomic systems is dealt with by Sommerfeld, *Wave Mechanics*, p. 16 (1930); Ruark and Urey, *Atoms, Molecules and Quanta*, p. 437 (1930).

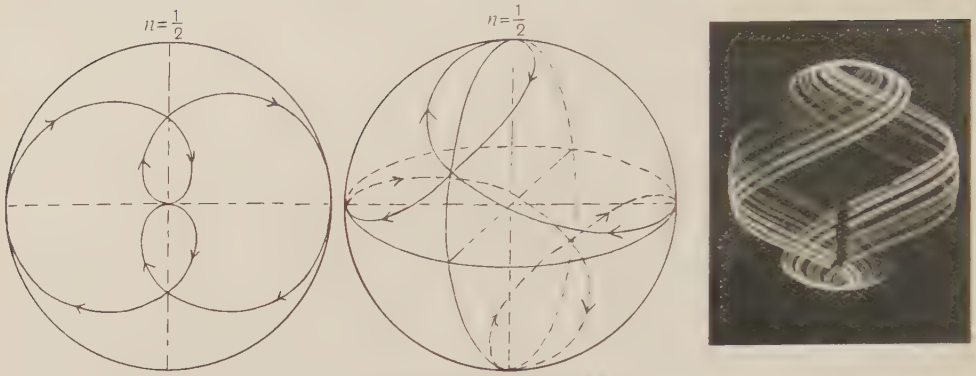


Figure 2 (*b*). $r=a\sin\frac{1}{2}\phi$.

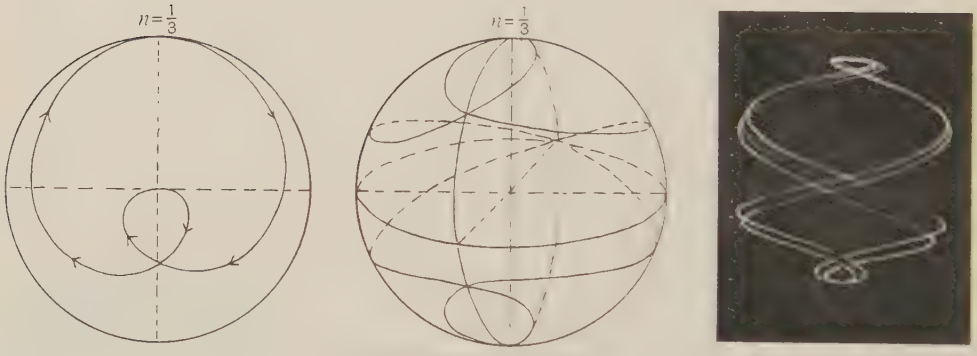


Figure 2 (*c*)*. $r=a\sin\frac{1}{3}\phi$.

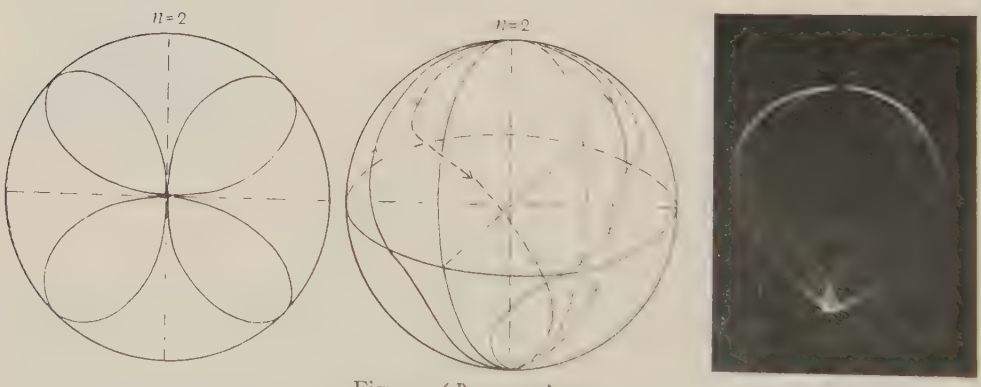


Figure 2 (*d*). $r=a\sin 2\phi$.

* Referred to its *node* as pole, this equation takes the form of the limaçon $r=\frac{1}{2}-\sin\phi$.

represents the successful attempt after some failures. It consists of a light arm B, 5 in. long, weighing only 1 gm., and pivoted at one end, with a small medical electric lamp P attached to the other end. The lamp weighs only 0·37 gm. It will be seen that P can move on the surface of an imaginary sphere of radius 5 in. The motions in co-latitude θ and azimuth ϕ are transmitted to the moving arm B by the two independently driven wheels R and S, R being fixed as shown to the hollow shaft and S to the flexible cable which runs through the hollow shaft, transmitting motion to the "fly-wheel" C and the connecting bevel gears. For success

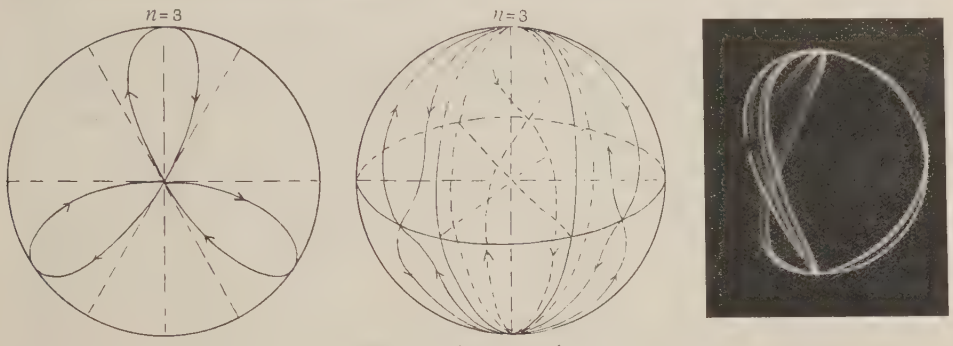


Figure 2 (*e*). $r = a \sin 3\phi$.

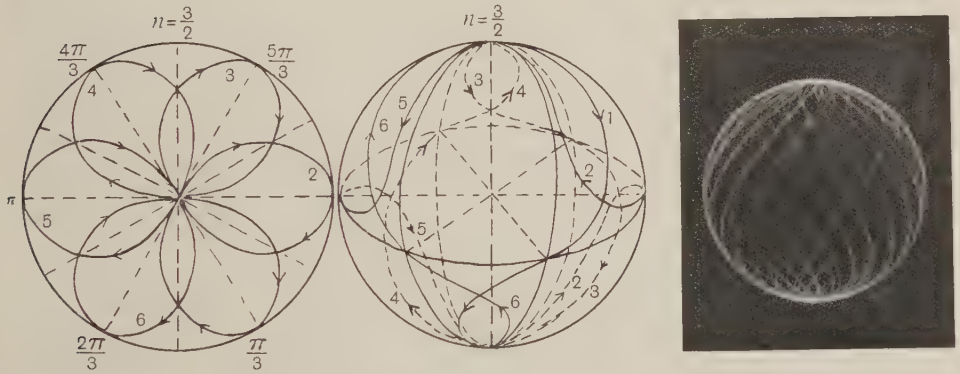


Figure 2 (*f*). $r = a \sin \frac{3}{2}\phi$.

in working (*a*) the bend in the upright hollow shaft must enable the lamp P to move unhindered to any part of the spherical surface; (*b*) the vertical axis of the shaft must be a free axis so as to avoid too much vibration (loads are suitably disposed along the axis to ensure this result); (*c*) the arm B and its attachments must have as small a moment of inertia I as possible in order to minimise the gyrostatic couple $I\dot{\theta}\dot{\phi}$ that comes into play during the motion. Finally (*d*) arrangements must be made for keeping the lamp alight during the whole of its motion. The method will be obvious from the figure; the current returns from the lamp through the metal apparatus itself.

I

Let ω_R, ω_S be the angular velocities of the wheels R and S respectively and in the same sense; then the angular velocity $\dot{\phi}$ of the lamp P in azimuth about the

ω_R, ω_S

vertical polar axis is ω_R, whilst the angular velocity $\dot{\theta}$ in latitude $= \omega_R - \omega_S$. The sign of ω must be changed if R and S rotate in opposite senses. If $\dot{\theta} = n\dot{\phi}$ as above, then the two possible methods of producing each figure are as shown in the table.

$n = \dot{\theta}/\dot{\phi}$	ω_R/ω_S
± 1	$0/\omega_R$ or $+2$
± 2	-1, $+3$
$\pm \frac{1}{4}$	$+\frac{3}{4}$, $+\frac{5}{4}$
$\pm \frac{3}{4}$	$\pm \frac{1}{4}$, $\pm \frac{7}{4}$

The figure seen depends upon the phenomenon of persistence of vision. The precession effect which occurs in azimuth if the ratio of $\dot{\theta}$ to $\dot{\phi}$ cannot be expressed as a commensurable number is shown in some of the accompanying photographs, which represent examples of the figures actually observed.

We wish to thank Mr F. Staley and Mr H. G. Bell for their kind assistance.

519.281.2

NOTES ON THE METHOD OF LEAST SQUARES

By Sir A. S. EDDINGTON, F.R.S.

Received November 10, 1932. *Read January* 20, 1933.

ABSTRACT. In inferring the value of a physical quantity x from observations some risk must be accepted. It is therefore presumed that the investigator has made up his mind how much risk he will take, and desires the closest possible limits—the narrowest range of values of x—that he can adopt without exceeding this risk. The aim is to furnish a concise treatment of combination of observations on this basis, which it is hoped may be found useful in clearing up common misconceptions. Stress is laid on the fact that the method of least squares is justified without the assumption of a Gaussian error law of the observations. Most of the paper deals with quite elementary points, but it ends with a discussion of more difficult questions which arise in inferring the mean square error of observation from the residuals.

§ 1. INTRODUCTION

The main purpose of this paper is to controvert two rather prevalent ideas: (1) that the method of least squares is only justified if the errors of observation have a Gaussian distribution, and (2) that the orthodox theory of errors does not sanction any of the other methods of combining observations that are often used in practice. I desire also that the theory should be presented in a way which does not introduce inverse probability. I realize that others have put forward similar or partly similar views, and probably all the substantial parts of this paper have been given at one time or another by earlier writers; nevertheless I hope that the presentation taken as a whole will show the subject in a new light to some readers*.

I think that the trouble arises because text-book writers often take it for granted that the object of the theory of combination of observations is to find *the most probable value* of the quantity x that is under discussion. This is a strange obsession, because it is difficult to imagine any circumstances in which it could be of interest to a scientist to know the most probable value of x. We must therefore insist at the outset that the purpose of the theory is to assist scientific investigation and not to answer mathematical conundrums.

What the physicist desires is *the most accurate determination* of x, in a sense which will be defined precisely in § 2. Further, although the superlative *most* may represent one aim, the more usual aim corresponds to the comparative *more*; we wish to judge of any proposed method whether it is more accurate, and how much more accurate it is, than another. No practical physicist has ever set himself to

* The main purpose is accomplished in §§ 1–6; §§ 7–10 concern those minutiae, beloved of the mathematical theorist, which cannot be wholly omitted.

make the most accurate determination conceivable of a physical constant; he usually wishes to make a more accurate determination than his predecessors; sometimes his intention is merely to obtain a determination which is *accurate enough*, i.e. which reaches some specified degree of accuracy. The calculator may well have similar aims.

§ 2. THE "MOST ACCURATE DETERMINATION"

The meaning of "more accurate determination" is best explained by an example. Suppose that I have occasion to use Planck's constant, and I find in reference books two determinations given as

$$h \times 10^{27} = 6{\cdot}551 \pm {\cdot}013,$$
$$h \times 10^{27} = 6{\cdot}547 \pm {\cdot}008.$$

Assuming that these are to be taken at their face value, I choose the latter; but my choice is not in any way guided by the consideration that on the evidence 6·547 is more probable than 6·551. Scientific results are not used that way. If any conclusion is dependent on the choice, the first thing to bear in mind is that in all probability the true value is neither 6·547 nor 6·551. To stand a reasonable chance of being right I must employ a range of values. I therefore proceed as follows. I first make up my mind how much risk of error I am prepared to take—for with any ordinary law of errors some risk is inevitable. I decide, let us say, to take a risk of 1 in 10; this is probably not greater than the risk I habitually take in drawing inferences from observational evidence. The theory of errors tells me that for this risk I must allow $2\frac{1}{2}$ times the probable error. Accordingly I rely on the first determination as a definite assurance that the value is between 6·519 and 6·583, and the second that it is between 6·527 and 6·567. Naturally the latter is the more useful to me since it limits the value to a narrower range; it is a closer or "more accurate" determination, and will lead to sharper conclusions in the deductions that I am intending to make. Note that a "determination" is not a value, but a range of values that can be accepted with a given risk—though we often loosely apply the word to the mid-value of the range.

The accepted risk will vary according to the purpose for which the determination is to be used and the rashness of the user. A risk of $\frac{1}{6}$, corresponding to an allowance of double the probable error, is perhaps the most usual allowance for ordinary purposes; but a navigator determining his position at sea would not be satisfied to run his ship on the rocks once in six times. The practice of allowing merely the probable error as margin is not unknown—a tacit admission by the perpetrator that his habit is to tell the truth only on alternate occasions.

The "most accurate determination" should accordingly be defined by the criterion of *range per risk*. It would, however, be difficult to utilize the definition strictly in this form. It might happen that in comparing two methods of determination, one gave the more accurate determination (i.e. narrower range) when the accepted risk was small and the other the more accurate determination when the risk was large. The "most accurate determination" would then be ambiguous until

the accepted risk had been specified. This is scarcely likely to occur in practice for reasons explained below; but it makes it necessary to adopt a simplified (conventional) definition for theoretical purposes, which avoids all ambiguity. We adopt the mean square error of x as the best general criterion of the accuracy of x. The criterion of mean square error will be in any case a suitable compromise between criteria based on small risks and on large risks respectively; so that it has almost the same practical utility as the criterion of range per risk that we should ideally prefer. But there is actually a better justification for using it.

When a quantity x is deduced (by any method of solution) from observational data, it has a frequency curve of error; that is to say, considering all possible ways of incidence of accidental error in the observed quantities, the resulting errors in the deduced value of x follow a frequency curve. The frequency curve of error of x must not be confused with the frequency curve of error of observation. We shall show in § 4 that in all ordinary cases the error law of x is approximately Gaussian even when the error law of observation is very far from Gaussian.

Since then in practice we have to compare determinations of x which have similar frequency curves (all being approximately Gaussian, even when the errors of observation are non-Gaussian), the possible ambiguity does not arise. The curves differ only in linear spread; and the range for given risk (whether large or small) is simply proportional to the mean square error, which can therefore be used as an equivalent measure of accuracy.

Thus the aim of the theory of combination of observations is to devise a method of solution which shall give a value of x having the smallest possible mean square error; so that we may deduce the smallest possible range of values within which x can be asserted to lie, subject to the given risk of error that we have made up our minds to accept.

§ 3. METHOD OF SOLUTION

We consider the problem of finding the most accurate determination of x from a set of equations of condition

$\mu_r, r,$

$$a_r x + b_r y + c_r z = \mu_r \qquad (r = 1, 2, \ldots n) \qquad \ldots\ldots(1), \qquad a, b, c, \; x, y, z$$

where μ_r has an error derived from any symmetrical* error-distribution which may differ as widely as we please from the Gaussian law. The error laws for μ_1, μ_2, etc. need not be the same, but we assume that they all have the same mean square error ϵ; if not, the equations must first be multiplied by appropriate factors so as to make the mean square error of the right-hand side the same for all, as in the ordinary process of assigning weights. ϵ

Whatever method of finding a value of x we adopt, our value will depend on the measured quantities μ_r, so that x can be regarded as a function of the μ_r.

* Practical methods are generally designed to secure a symmetrical law, e.g. methods of reversal and differential measurements. But the restriction to symmetrical laws could be replaced by any other condition which excludes systematic error.

Assume for trial a linear function

$$x = \Sigma_r \lambda_r \mu_r \qquad \text{.....(2),}$$

λ where the λ_r are arbitrary coefficients. Then, on insertion of the value of μ_r from (1)

$$x = \Sigma\lambda_r a_r . x + \Sigma\lambda_r b_r . y + \Sigma\lambda_r c_r . z \qquad \text{.....(3).}$$

This is satisfied if

$$\Sigma\lambda_r a_r = 1, \quad \Sigma\lambda_r b_r = 0, \quad \Sigma\lambda_r c_r = 0 \qquad \text{.....(4).}$$

Any set of values of λ_r which satisfies (4) will give a "solution" $x = \Sigma\lambda_r \mu_r$. It is a solution because if the measured quantities μ_r are free from error it will give the correct value of x. But when the μ_r are affected by error, these different solutions may be of widely different accuracy and we have to determine that set of values of λ_r which gives the most accurate solution.

Before proceeding further the question may be raised whether the most accurate solution is necessarily included in the linear form (2). It is quite likely that for particular non-Gaussian error laws non-linear solutions may exist which are more accurate. But it is unlikely that any non-linear method would be seriously contemplated as practicable having regard to the immense labour involved and the rarity of the occasions which call for it. An exception should, however, be noticed. When there is only one unknown and the equations of condition become simple measurements of x, "fancy methods" can be used and may be more accurate; thus, if it is known that the error law is $e^{-k\epsilon}\, d\epsilon$, the median is more accurate than the arithmetic mean, at any rate for some risks. The median is an example of a solution not included in the form (2).

If the actual error of μ_r is ϵ_r, we have by (2)

$$\text{error of } x = \Sigma\lambda_r \epsilon_r \qquad \text{.....(5),}$$

and by a well-known elementary theorem (not assuming the Gaussian law)

$$\text{mean square error of } x = \sqrt{(\Sigma\lambda_r^2)} . \epsilon \qquad \text{.....(6),}$$

where ϵ is the mean square error of μ_r (the same for all the μ_r's as already stated). Thus the determination of x with the smallest mean square error to be obtained by linear methods is found by making $\Sigma\lambda_r^2$ a minimum, subject to the conditions (4) being satisfied. The minimum condition, together with the conditions given by differentiating (4), gives

$$\Sigma\lambda_r \delta\lambda_r = 0 \qquad \text{.....(7)}$$

for all variations $\delta\lambda_r$ which satisfy

$$\Sigma a_r \delta\lambda_r = 0, \quad \Sigma b_r \delta\lambda_r = 0, \quad \Sigma c_r \delta\lambda_r = 0 \qquad \text{.....(8).}$$

This is solved by the method of indeterminate multipliers. Suppose that

$$\lambda_r = a_r P + b_r Q + c_r R \qquad \text{.....(9),}$$

P, Q, R where P, Q, R are arbitrary coefficients independent of r. Then by (9)

$$\Sigma\,(\lambda_r - a_r P - b_r Q - c_r R)\,\delta\lambda_r = 0 \qquad \text{.....(10)}$$

for *all* variations $\delta\lambda_r$. But

$$\Sigma\,(a_r P + b_r Q + c_r R)\,\delta\lambda_r = 0 \qquad \text{.....(11)}$$

for those variations $\delta\lambda_r$ which satisfy (8). Adding (10) and (11) we see that $\Sigma\lambda_r\delta\lambda_r = 0$ for variations $\delta\lambda_r$ which satisfy (8). Accordingly by taking λ_r of the form (9) the condition for a *minimum* is automatically satisfied, and it only remains to choose P, Q, R so that the conditions for a *solution* (4) are satisfied.

Substituting (9) in (4), we have

$$\left.\begin{aligned} \Sigma a_r^2 P + \Sigma a_r b_r Q + \Sigma a_r c_r R &= 1 \\ \Sigma a_r b_r P + \Sigma b_r^2 Q + \Sigma b_r c_r R &= 0 \\ \Sigma a_r c_r P + \Sigma b_r c_r Q + \Sigma c_r^2 R &= 0 \end{aligned}\right\} \quad \text{.....(12),}$$

so that P, Q, R can be determined. Write the determinant of the coefficients of P, Q, R on the left of (12) in the standard form

$$\Delta = \begin{vmatrix} a & h & g \\ h & b & f \\ g & f & c \end{vmatrix} \qquad \Delta$$

and use capital letters for the minors of the small letters. The solution of (12) by the method of determinants is A, H, G

$$P = A/\Delta, \quad Q = H/\Delta, \quad R = G/\Delta \qquad \text{.....(13).}$$

Hence by (9)

$$\lambda_r = (a_r A + b_r H + c_r G)/\Delta \qquad \text{.....(14),}$$

and by (2)

$$\begin{aligned} x &= (\Sigma a_r \mu_r . A + \Sigma b_r \mu_r . H + \Sigma c_r \mu_r . G)/\Delta \\ &= \begin{vmatrix} \Sigma a_r \mu_r & \Sigma a_r b_r & \Sigma a_r c_r \\ \Sigma b_r \mu_r & \Sigma b_r^2 & \Sigma b_r c_r \\ \Sigma c_r \mu_r & \Sigma b_r c_r & \Sigma c_r^2 \end{vmatrix} \div \Delta \qquad \text{.....(15).} \end{aligned}$$

This then is the value of x with the smallest mean square error obtainable by linear methods. It is identical with the value obtained by solving the normal equations of least squares*.

Further by (14)

$$\begin{aligned} \Sigma\lambda_r^2 &= [(\Sigma a_r^2 A + \Sigma a_r b_r H + \Sigma a_r c_r G)\, A + (\Sigma a_r b_r A + \Sigma b_r^2 H + \Sigma b_r c_r G)\, H \\ &\qquad + (\Sigma a_r c_r A + \Sigma b_r c_r H + \Sigma c_r^2 G)\, G] \div \Delta^2 \\ &= [(aA + hH + gG)\, A + (hA + bH + fG)\, H + (gA + fH + cG)\, G] \div \Delta^2 \\ &= (\Delta A + 0 + 0)/\Delta^2 = A/\Delta = P. \end{aligned}$$

Hence by (6)

$$\text{mean square error of } x = \epsilon\sqrt{(A/\Delta)} \qquad \text{.....(16).}$$

This also agrees with the result of the method of least squares; in fact the coefficient A/Δ, or P, is usually found in practice by solving the equations (12).

Our results (15) and (16) thus justify the method of least squares without postulating a Gaussian error law, provided that it is understood that the method is

* So far as the mathematics are concerned, I have utilized the method of Laplace (1811). I do not think it can be appreciated when presented apart from the argument to which I have here adapted it.

not concerned with "most probable values." Its achievement is that (1) it gives a solution, i.e. a value which would be correct if the observations were exact; (2) it gives the mean square error of the result so that we know how wide a margin about the solution value we must allow; (3) no other linear method (and therefore no other method which we should be likely to contemplate employing) gives so small a mean square error and hence so narrow a range of x that can be trusted with an assigned risk. The conceivable existence of other methods (non-linear) giving still greater accuracy need not unduly trouble us; in practice it is more essential to know our accuracy than to crow about it.

§ 4. FREQUENCY LAW OF x

By (5) the error of x is made up of n terms each involving an independent error, and in all fair applications of the method of least squares n is somewhat large. Although some of the coefficients λ may turn out to be comparatively small, a good proportion of them should be of about equal order of magnitude. For if two or three of them outweigh all the others, so that they are practically the only ones that count in the formula for x, it means that the investigator has included a mass of observational material which has practically no bearing on the determination of x, though it concerns the other unknowns in the equations of condition. He is really relying on two or three measures and not on n measures as he supposes. Thus, apart from obviously inappropriate applications of the method, we find that the error of x is the resultant of a considerable number of independent errors of more or less equal importance. It is well known that in these conditions the resultant error is approximately Gaussian even when the contributing errors (errors of observation) are non-Gaussian. The number of contributing errors need not be unduly large; 6 or 8 will give a good Gaussian curve

It is interesting to note the greater security and rigour attained when we realize that any assumption of Gaussian distribution that we require refers to the error of x and not to the errors of observation. Apart from empirical evidence, the only "proof" of the Gaussian law of errors of observation rests on the assumption that the error of observation is the resultant of many small independent errors, objective or subjective*. The assumption that the *observer*, in making his error, combines a lot of little independent errors may perhaps be defended, but it offers a favourite target to the critic; it is much more satisfactory to be able to substitute the assumption that the *calculator*, in reaching his result, combines a lot of little independent errors, for we can always examine the actual coefficients λ_r to see whether this is true, and confound (or justify) the critic.

§ 5. NON-GAUSSIAN ERRORS

It is well worth while to place our theory of combination of observations on a basis which is not restricted to observations obeying the Gaussian law, for exceptions to this law are not entirely trivial. It is well known that there is a widely prevalent

* The proof from "the principle of the arithmetic mean," which is still sometimes quoted, is altogether fallacious.

cause tending to make observational errors follow the Gaussian law, viz. the combination of a large number of sources of small errors; but it is not so well known that there is a widely prevalent cause tending to bring about systematic deviation from the Gaussian law, viz. heterogeneity of observational material. This last cause operates when the data are supplied by a number of observers of different skill or by the same observer working under conditions which have varied. It gives an error law with an excess of large and small errors and a defect of intermediate errors, in comparison with the Gaussian distribution; analytically, the ratio of the fourth moment to the second moment is greater than in the Gaussian law.

For a Gaussian distribution the mean square error ϵ_2 and the mean fourth power error ϵ_4 are given in terms of the measure of precision h by ϵ_2, ϵ_4, h

$$\epsilon_2{}^2 = 1/2h^2, \quad \epsilon_4{}^4 = 3/4h^4 \qquad \text{.....(17).}$$

Consider batches of n_1, n_2, n_3, ... Gaussian errors with respective measures of precision h_1, h_2, h_3, ... Then for the mixture n

$$\epsilon_2{}^2 = \Sigma n_r/2h_r{}^2 \div \Sigma n_r,$$
$$\epsilon_4{}^4 = \Sigma 3n_r/4h_r{}^4 \div \Sigma n_r.$$

Hence

$$\tfrac{4}{3}(\Sigma n_r)^2(\epsilon_4{}^4 - 3\epsilon_2{}^4) = (\Sigma n_r)(\Sigma n_r/h_r{}^4) - (\Sigma n_r/h_r{}^2)^2$$
$$= \sum_{rs} n_r n_s \left(\frac{1}{h_r{}^4} + \frac{1}{h_s{}^4} - \frac{2}{h_r{}^2 h_s{}^2}\right) = \sum_{rs} n_r n_s \left(\frac{1}{h_r{}^2} - \frac{1}{h_s{}^2}\right)^2 \qquad \text{.....(18),}$$

which is a positive quantity, vanishing when the h's are all equal. Thus $\epsilon_4{}^4 = 3\epsilon_2{}^4$ for the Gaussian law, but $\epsilon_4{}^4 > 3\epsilon_2{}^4$ when batches of observations of different degrees of precision are pooled—as must usually happen in practice. It might perhaps have been thought that if the precision h were itself randomly distributed the Gaussian law would be recovered, but the formula shows that this does not happen.

§ 6. SIMPLIFIED METHODS OF SOLUTION

The investigator is free to choose for his solution any other set of values of λ_r which satisfies (4), and he often does so to save labour of calculation. His mean square error is given by (6), and his solution is a valid determination of x having the degree of accuracy thus assigned. It is necessarily less accurate than the least squares solution; but to tell him that he would have found a more accurate value by performing, say, three times as much numerical calculation is as impertinent as to tell an experimenter that he would have obtained a more accurate result by making three times as many experiments. It is not a crime to fail to reach the highest possible accuracy; it is a crime to fail to reach the accuracy which you claim.

Perhaps the commonest case in which a simple alternative to the method of least squares can safely be recommended is typified by the determination of the period T of a pendulum by observing the times t_r of n consecutive passages. The equations of condition are T, t, n

$$a + rT = t_r \qquad (r = 0, 1, 2, \ldots n - 1).$$

One linear solution is obtained by comparing the mean of the first m times $(t_0, \ldots t_{m-1})$ with the mean of the last m times $(t_{n-m}, \ldots t_{n-1})$. The difference gives $(n - m)\ T$ with a mean square error $\epsilon\sqrt{2}\sqrt{m}$, so that T is determined with a mean square error

$$\frac{\epsilon\sqrt{2}}{(n-m)\sqrt{m}} \qquad \ldots\ldots(19).$$

The weight of the determination (proportional to the inverse square of the mean square error) is therefore $\frac{1}{2}m\,(n-m)^2$. This is a maximum when $m = \frac{1}{3}n$. Hence the most accurate solution of this kind is obtained by comparing the mean of the first third against the mean of the last third of the passage times, the determination being then of weight $2n^3/27$.

The weight of the determination of the value of T by the method of least squares is, of course, rather greater; it is found to be $n\,(n-1)\,(n-2)\ 12$. Thus when n is large the weights of the two solutions are in the ratio $8 : 9$. It is seldom that the difficulty of obtaining additional observational material is so great as to warrant the trouble of a least-squares solution for so slight a gain. The "method of thirds" is generally useful when we have to determine a slope or a linear scale-constant from more or less uniformly spread data.

§ 7. INVERSE PROBABILITY

The ordinary presentation of the method of least squares as a method of finding the most probable value of x involves an appeal to inverse probability which is difficult to justify. The argument runs as follows. If the true values of the unknowns are x', y', z', we find (from the residuals of the equations of condition) that we must have made errors $\epsilon_1', \epsilon_2', \ldots \epsilon_n'$ in our observations; if on the other hand the true values are x'', y'', z'', our observational errors must have been $\epsilon_1'', \epsilon_2'', \ldots \epsilon_n''$. We find further that $\epsilon_1', \epsilon_2', \ldots \epsilon_n'$ is a more probable set of errors than $\epsilon_1'', \epsilon_2'', \ldots \epsilon_n''$. We then introduce inverse probability, and argue that since we are more likely to have made the errors $\epsilon_1', \epsilon_2', \ldots \epsilon_n'$ than the errors $\epsilon_1'', \epsilon_2'', \ldots \epsilon_n''$, it is more likely that the true values are x', y', z' than that they are x'', y'', z''.

The argument that, because the consequences of hypothesis A turn out to be more probable than the consequences of hypothesis B, therefore hypothesis A is more probable than hypothesis B, cannot be accepted without qualification. Having tossed a penny 5 times in succession we observe that it has come down heads every time; this occurrence is much more probable on the hypothesis that it is a double-headed penny than on the hypothesis that it is an ordinary penny, but it does not follow that it is more probably a double-headed penny than an ordinary penny. It is true that in rejecting the argument we are guided not only by the ostensible data of the problem but by private knowledge that double-headed pennies are rare; but the real point is that questions of inverse probability are of such a nature that they compel us to introduce private knowledge or prejudice—that is to say, a priori expectation—since otherwise there is no answer to them. The objection to inverse probability is not so much that it is inadmissible as that it is silly.

The inverse-probability argument in the theory of least squares is valid only on the understanding that a priori all values of x, y, z are equally likely. Starting on this basis, we can then introduce the additional information supplied by our observations (which, of course, modifies the a priori expectation) and deduce rigorously that x', y', z' is more probable than x'', y'', z''. But we are saved any debate as to whether this hypothesis of equal a priori probability of all values of x is an admissible one, because it is easily seen that it cannot be generally true. Certain methods of determining the charge of an electron provide linear equations of condition in which the unknown quantity x is e; others yield equations of condition in which x is e^2. If all values of e are equally probable a priori, then all values of e^2 cannot be equally probable a priori; so that the hypothesis is untrue in one case. In practice the range of values of e that need seriously be considered is so small that it makes no appreciable difference whether the hypothesis of equal a priori probability refers to e or e^2 or any function of e within reason. The difficulty may not be a practical one, but it shows that the assumption underlying the use of inverse probability in the usual theory of least squares has no axiomatic basis.

This difficulty in the usual theory of least squares arises because it mixes together two distinct subjects, viz. the combination of observations and the philosophy of inference from observation. The presentation of the theory here adopted avoids the difficulty entirely because it sticks strictly to the combination of observations. It gives us a solution, say $x = x_0$, with a mean square error μ. We have first the satisfaction of knowing that it was not within our power to reduce μ by adopting any other practicable method. Secondly we may treat our result and draw just the same inferences as if we had found x_0 by a single direct observation made with mean square error μ*. Effectively then we have replaced a mass of complicated observational material by the *best possible equivalent single direct observation* of the quantity x that we are seeking. The task of the theory of the combination of observations ends with this reduction. Anything further that we may add belongs to the philosophy of inference from an observation. This resolves itself into a simple question—How are we justified in making any inference from an observation which is subject to error?

x_0, μ

The answer that I have adopted is as follows. We can never be sure of our inferences; therefore we have to aim at a system of inference that gives conclusons of which in the long run only $1/q$ are wrong. Accepted observational knowledge of the universe is a function of q—a series of maps becoming more and more detailed as q decreases. The maps start (at $q = \infty$) with one which is entirely trustworthy and entirely blank; and they end (at $q = 0$) with one loaded with detail of which only an infinitesimal proportion is correct. How a philosopher would use these maps I have no idea. I only affirm that some of the intermediate maps (say q 5

q

* Except that the combination of observations has one advantage over the single observation. In drawing inferences from a single observation it is usually assumed rather hypothetically that the error law is Gaussian, but as was shown in § 4 we have much greater security that the error law of x_0 is approximately Gaussian.

to $q = 20$) can be of considerable assistance to a being who has to find his way about in the universe he lives in.

The answer given by some is that we infer (from the observation) an inverse probability of true values of x. But then, what are we to infer from the inverse probability?

§ 8. DETERMINATION OF THE MEAN SQUARE ERROR

After a least-squares solution has been made, it is usual to compute from the residuals the mean square error of an observation. I add some remarks on this part of the problem, but limit them to errors obeying the Gaussian law.

μ If μ is the mean square error computed directly from a particular sample of n errors, it can be shown that the frequency distribution of μ is

$$\mu^{n-1} e^{-n\mu^2/2\epsilon^2} d\mu \qquad \text{.....(20),}$$

where ϵ is the true mean square error. The same distribution (20) is obtained when m μ is computed from a sample of $n - m$ *residuals*, m being the number of unknowns $\bar{\mu}$ in the equations of condition. Hence the mean value $\bar{\mu}$ of μ found from an unlimited number of such samples is given by

$$\bar{\mu} = \int_{-\infty}^{\infty} \mu^n e^{-n\mu^2/2\epsilon^2} d\mu \div \int_{-\infty}^{\infty} \mu^{n-1} e^{-n\mu^2/2\epsilon^2} d\mu$$

$$= p\epsilon \qquad \text{.....(21),}$$

p where

$$p = \sqrt{\frac{2}{n}} \cdot \frac{\Gamma\left(\frac{1}{2}n + \frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}n\right)} < 1 \qquad \text{.....(22).}$$

The question is sometimes raised whether μ should be regarded as a determination of ϵ or of $p\epsilon$. It can be shown that its mean square deviation from ϵ is

$$\epsilon\sqrt{\{2(1-p)\}} \qquad \text{.....(23),}$$

and from $p\epsilon$ is

$$\epsilon\sqrt{(1-p^2)} \qquad \text{.....(24).}$$

Since $p < 1$, the latter is the smaller. Thus absolutely μ is a closer "shot" at $p\epsilon$ than at ϵ. But since $p\epsilon < \epsilon$, a given mean square error of $p\epsilon$ is equivalent to a larger mean square error in ϵ; and it turns out that *proportionately* μ is a closer shot at ϵ. For if we set $p\epsilon = \mu$, we determine ϵ with a mean square error

$$\frac{\epsilon\sqrt{(1-p^2)}}{p},$$

whereas if we set ϵ equal to μ, we determine ϵ with the mean square error (23) which is easily shown to be smaller. Thus it is better to regard μ as a determination of ϵ, not $p\epsilon$.

This point is in any case rather a trivial one, and it scarcely arises when we remember that for any useful application what is needed is not a value of ϵ but a range of values which can be trusted with given risk.

§ 9. INFERENCE OF ϵ FROM μ

The problem of inference from μ is in principle the same as the problem of inference from a solution value of x already discussed. The theoretical calculation of μ reduces the mass of observational material to a single shot at ϵ. We can treat μ just as though it were a direct observation of ϵ subject to observational error. The error law is not Gaussian and is in fact slightly skew; but it is given explicitly in formula (20), so that we can find from it the range corresponding to any risk.

A practical difference, however, arises from the fact that in a determination of x the error curve of x is supposed to be independent of the value of x, whereas in a determination of ϵ the scale of the error curve of ϵ varies proportionately to ϵ. We have therefore to proceed in a rather indirect way if we are to avoid introducing inverse probability. Suppose as before that we accept a risk of $\frac{1}{10}$. From (20) we can calculate factors k, k' (functions of n) such that there is a $\frac{9}{10}$ chance that μ is between the limits ϵ/k and $k'\epsilon$. Hence having observed a particular μ, we can infer that either ϵ is between the limits μk and μ/k', or that a $\frac{1}{10}$ chance has come off. Hence if we make a practice of inferring from a determination μ that ϵ is between μk and μ/k', we shall only be wrong on those occasions on which a $\frac{1}{10}$ chance has come off—in the long run once in 10 times. Thus μk to μ/k' is the range which we can adopt with the accepted risk. k, k'

There is a rather subtle distinction between what we have just said and the statement that the probability of ϵ being between μk and μ/k' is $\frac{9}{10}$. The latter would be a statement of inverse probability. Reverting to the example of a penny turning up heads on the first 5 throws, we can certainly infer that either the penny is double-headed or a $\frac{1}{32}$ chance has come off. But I think no one would say that the probability of its being double-headed is $\frac{31}{32}$; the probability is simply non-existent until some information as to the a priori frequency of double-headed pennies is supplied. All the same, if we pursue penny-tossing as an occupation, and habitually infer when the first 5 throws turn up heads that the penny is double-headed, and infer when a tail appears in the first 5 throws that it is not double-headed, we shall only be wrong when a $\frac{1}{32}$ chance has come off, and perhaps not then if double-headed pennies exist. Thus in the long run we shall not be wrong more than once in 32 times. The distinction is between the probability of a particular affirmation and the probability of our making a right decision by the method which led to the affirmation.

To sum up: when ϵ is known we know the frequency distribution of μ, but we cannot solve the inverse problem of finding the frequency distribution of ϵ when μ is known. The latter is not required because we can infer directly from μ a range of ϵ, subject to the inevitable limitation that a specified proportion of our inferences by this method will be wrong.

§ 10. COMBINED INVESTIGATION OF x AND ϵ

In writing down the range of x which can be accepted with given risk, we use the mean square error of x, viz. $\epsilon\sqrt{(A\,\Delta)}$, which is known in terms of the mean square error of an observation ϵ. But if ϵ itself is inaccurately known and can be specified only as a range (subject to risk), a complicated problem arises as to how to combine risks in ϵ and risks in x. Owing to the fact that a rather large number of residuals are needed in order to determine ϵ at all sharply, this is not always a pedantic objection. There is therefore considerable interest in the theory (now often preferred) by which the frequency distribution of x is found directly in terms of the observed mean square error μ instead of in terms of the true mean square error ϵ whose value is uncertain. I think, however, that this theory has not been extended to non-Gaussian errors of observation.

This undoubtedly is the more logical procedure when x and ϵ are determined *from the same data*, as is often the case. But it would not be quite fair to the ordinary theory of least squares to regard it as superseded by this modern refinement. More especially in those problems of astronomy and geodesy which form the traditional field of application of the theory of errors, the determination of ϵ rests on much more extensive data than the determination of x; so that the determination of ϵ and the determination of x are separate problems as here assumed. From the practical point of view, adequacy of data is more important than mathematical polish. For example, astronomers determining a hundred or more stellar parallaxes used to calculate the probable error of each parallax from its own residuals; it has been realized that this is very wasteful of accuracy, and nowadays the probable errors of the separate parallaxes are based on a common value of ϵ obtained from the residuals of the whole series. Thus the determination of ϵ employs more than a hundred times as many data as the determination of an x.

I have therefore felt justified in ruling the rigorous combined treatment of ϵ and x to be outside the scope of this paper, regarding it as an incidental rather than a basic problem in the theory of least squares.

DISCUSSION

Dr N. R. Campbell. I agree so entirely with the author's conclusions that I hesitate to criticize his arguments. But the following considerations seem to support rather than to contradict them.

(1) The author says rightly that the physical problem is always to select a range of values. He implies, rightly again, that, in using the selected range to test the truth of a physical proposition, it is always assumed tacitly that all values within this range are equally probable. If so, the application to physical problems of a theory whose aim is "to devise a method...which shall give *a* value of x..." (uniquely distinguished) must involve logical inconsistency. Such a theory cannot, therefore, be used to *prove* anything physically important; it can only *suggest* something to be established by other means.

(2) The assumption that probability is determined by the mean square error is convenient and often legitimate. But there is no evidence that it is accurately true. Consequently it may be expected sometimes to exclude permissible methods and sometimes to justify impermissible methods. Thus the "method of thirds" must be wrong; for the best method of using a group of equally trustworthy observations can never be to ignore completely one-third of them. This is important, because the theory of errors is the last surviving stronghold of those who would reject plain fact and common sense in favour of remote deductions from unverifiable guesses, having no merit other than mathematical tractability.

(3) Six or eight observations will *not* give a good Gaussian curve; six or eight millions will not do so. For the Gaussian curve is continuous, while the frequency curve of observations must be discontinuous, because the number of alternative readings on any instrument whatever must be finite. Again, the assumption of continuity is sometimes permissible; but, once more, experience shows that it is likely to be used most confidently when the evidence against it is most overwhelming. (See R. T. Birge's recent papers.)

Inst.-Capt. T. Y. Baker. May I plead for a more explicit statement on the part of authors of scientific papers as to what they mean when they sum up the result of a series of measurements as $a \pm b$? For I am sure that the conventions that underlie, or ought to underlie, this summarization are often unknown to the reader and sometimes to the author. I believe that a is, conventionally, the algebraic mean of the n values and b is $\mu/\sqrt{n}$, where μ is the mean square error. That is to say, if r is a typical residual, $\mu = \sqrt{(\Sigma r^2/n)}$, so that $b = \sqrt{(\Sigma r^2)}/n$. I am pretty certain that b does *not* always have this value and may on occasion stand for probable error, or mean error, or mean square error, or even for some multiple of one of these. Doubtless to anyone continually engaged on work of a statistical nature, and who is frequently applying the methods of the theory of errors, no ambiguity is likely to arise, but so many people who read the journals of scientific societies have no intimate knowledge of that branch of mathematics. The risk of the expression $a \pm b$ being misunderstood is so great that I feel sure an author would be doing a great kindness to his readers if he would always state, quite clearly, exactly what he means when he puts down an expression of this character. Let me give an example of the way in which a misunderstanding may arise. In a dimensioned engineering drawing $a \pm b$ is very frequently used. The nominal dimension is a; b is the tolerance either way. An engineer looking up the value of Planck's constant, mentioned in § 2, might, quite naturally, infer that $h \times 10^{27}$ might lie anywhere between 6·538 and 6·564, according to one authority, and between 6·539 and 6·555 according to another, whereas the limits that connote practical certainty are wider.

One cannot have the whole of a series of measurements printed *in extenso*, and if a convention as to the meaning of b is required I suggest that the mean error of a single observation has much to recommend it. In the first place, the mean error involves less labour in its computation and furnishes a check, in the equality of the sums of the positive and negative residuals, on the accuracy of the arithmetic. With

the very limited number of observations to which the Gaussian law of errors is so often applied, the modulus of the curve may, for all practical purposes, be derived just as well from the mean error as from the mean square error. In the second place, the mean error gives simple factors for the ranges corresponding to risks of 1 in 10, 1 in 100 and 1 in 1000. They are respectively 2, 3 and 4 times the mean error. Thus, expressed in this manner, $a - b$ indicates immediately that there is one chance in 10 that the value of a single observation lies outside the range $a - 2b$ to $a + 2b$, one chance in 100 that it lies outside the range $a - 3b$ to $a + 3b$, and one chance in 1000 that it lies outside the range $a - 4b$ to $a + 4b$. These numbers are only approximate; more accurately they should be 2·06, 3·22 and 4·14, but the round figures are generally good enough for practical purposes.

Finally I would urge physicists not only to explain at the head of every paper what they mean by $a \pm b$, but also to make a point of giving the number of observations in each case.

Mr J. Guild. This paper is one which will be read with great interest by experimental physicists: it would undoubtedly have given rise to a prolonged and useful discussion had it been presented otherwise than as one item in a rather congested programme.

It is gratifying to find such an eminent mathematician as the author agreeing that there is no validity in the Gaussian law as applied to errors of observation. Practical physicists have challenged it before, on the ground that it bears no relation whatever to the circumstances in which errors of observation arise, but little attention appears to have been paid to them. N. R. Campbell* rejects the Gaussian theory of error "without qualification and with the utmost possible emphasis; and with it... all theoretical grounds for adopting the rules that are based on it." His reasons for doing so must appeal to all practising physicists as in complete accordance with the facts of their own experience if they have troubled to give any thought to the matter at all. Any theory based on an assumed law for the distribution of observational errors must be ruled out of court.

I am glad to note, also, that the author effectively disposes of the meaningless old bogey, "most probable value," but would have liked him to make a similar clearance of the term "probable error," which is a source of great confusion and seems to me to have no more justification than the other.

Having discarded the Gaussian law as applied to observational errors, the author reintroduces it as applied to the error of x, where x is the value of the thing measured, deduced from a series of observations each subject to a small independent error, and he re-establishes the method of least squares on this basis. It would appear that under circumstances, if such ever exist, in which we can agree in advance that the observational errors are "independent" no exception can be taken to this process. I would not like to say that such circumstances never exist, but I would hesitate equally to admit that I had ever met them. On the other hand, one can assert with certainty that in at least many, if not all, of the cases which arise in laboratory

* *Measurement and Calculations*, p. 162.

physics there is no such thing as an entirely independent error. Errors of the type we are here considering, namely those which produce a spread in the observations, arise from failure in one or all (usually all) parts, human and mechanical, of the experimental system to behave in an absolutely identical manner under immediately identical circumstances. Past history affects the behaviour not only of the human observer but of every constituent part of the apparatus. I suppose, however, that independence of the errors, as such, does not enter into the author's argument, provided that, whether independent or not, they give rise to an approximately Gaussian distribution in the resulting values of x; and the advantage of this criterion, as the author points out, is that the computer can always see whether it is satisfied or not in any particular case.

I am not convinced, however, that even then Campbell's case against the Gaussian method of reduction has been invalidated. It is generally admitted, except possibly by rabid statisticians, that no method of reduction of experimental observations can give an answer which has any significance within limits which are *very* much closer than the uncertainty of the individual observations. This is certainly the case in all laboratory work, where time and other considerations prevent the enormous series of observations, with which the statistician loves to work, from being made; and no competent experimenter would ever attach significance to the *difference* between the results deduced from a given series of observations by different legitimate methods; he would always use the method which involves the least arithmetic, such as taking the mean value of the relevant variable. If the difference between this and any other feasible result based on the observations is comparable with the range per risk at which he is aiming, his time will be much better spent in improving his technique, so as to get better observations, than in making lengthy calculations.

With regard to the evaluation of range per risk as a criterion of the reliability of a determination, I submit that "probable error," or any other function of the non-systematic errors, is entirely valueless. The risk of the true result being outside any given limits is surely as much dependent on systematic error as on the spread of the observations. No theory of errors which aims at obtaining formulae can take systematic error into account. The mathematician generally assumes that it is the business of the experimenter to eliminate all systematic errors. So it is, and he does so as completely as he can; but when the precautions suggested by his experience and permitted by his available equipment are exhausted some sources of systematic error still remain and, from the point of view of the experimenter who is doing his best to measure something, such unsuspected errors are as much accidental as those which merely affect the precision of repetition. They are a much more serious problem to him, or to anyone else who wants to estimate the accuracy of the result. Criteria of precision may have been of great importance in the early days of experimental science when relatively crude and insensitive instruments had to be used, but nowadays it is nearly always possible in any type of measurement to get more precision than has any significance in the final result. The elusive fiend—systematic error—is the crucial thing in modern laboratory work, and the

man who estimates the value of his own determinations by any function of the precision is merely singing himself to sleep with a pretty formula. As the result of twenty years' experience in a laboratory where accuracy is the primary object in all the work carried out, I would attach absolutely no importance whatever to $\pm$ limits, based on root-mean-square error or any other function of non-systematic error, placed after results obtained either by myself or anyone else. If I were faced with the particular problem put by the author at the beginning of paragraph 2, I would take no notice of the $\pm$ 0·008 or $\pm$ 0·013. I would first read carefully through the papers of both men and form some idea for myself of the relative merits of their technique and of their experience in work of this kind. If they happened to be well-known men I would supplement my own judgment by consideration of the instances known to me in which their work had been contradicted or corroborated by others of similar status, and from these various considerations I would acquire a greater amount of faith in one result than in the other. I might, of course, be wrong in any particular case, but in the long run I should be more often right than wrong because I am taking into account the really relevant factors, and this is what no formula based on mere precision of observations can ever do.

Mr T. Smith. Sir Arthur Eddington is very closely associated with a science in which very effective use has been made of the theory of errors. His views will interest even those who consider this theory of no importance so far as their own branch of work is concerned, and in this number are several members of the N.P.L. staff. Assuming, however, that there are types of measurement carried out at the N.P.L., or similar standardizing institutions, where the theory can properly be applied, what does Sir Arthur consider a reasonable risk for them to run in giving results? Since these results are given for use outside the Laboratory, the risk is not a matter which concerns the staff alone.

I should also like to ask Sir Arthur to give his views on the application of the theory to the following problem. Two potential differences, each in the neighbourhood of 100 volts, are to be compared, and the comparison is to be correct (the risk, say, being that permissible for a standardizing laboratory to run) to 0·1 volt, so that the tolerance is represented by $\pm$ 0·05 volt. The only instrument available at the moment for the comparison is a correctly graduated voltmeter, which, however, has the property that the pointer will not move from a given position unless the voltage changes by at least 1 volt, and that the pointer, owing to grit, to inertia, or other causes, may come to rest at any place within $\frac{1}{2}$ volt of the applied potential difference. Stated in this form the problem may seem frivolous, but it represents a class of real problems on which divergent views are held, and some of us would like to hear the author's.

Author's reply. I am very pleased that the paper should have drawn so interesting a discussion. In reply to Dr Campbell's points as numbered: (1) I said, the aim of the theory is "to devise a method...which shall give a value of x...so that we may deduce a range...." By omitting the last words Dr Campbell makes it appear that the final result is *a* value instead of a range, contrary to the whole

tenor of the paper. I do not anywhere assume or imply that values within the range are equally probable; in general, that is far from true. Anything less than the whole range corresponds to a risk greater than we can accept, so that no question of subdividing the range arises. (2) As Mr Guild points out, there is no object in reducing the accidental error much below the unavoidable systematic errors; so when the method of thirds reaches this limit of accuracy there can be no justification for preferring more complicated methods. In any case I object to describing a method as "wrong" merely because a more accurate result could have been obtained by additional labour. By that standard all quantitative experimental results are wrong. (3) I did not say that 6 or 8 observations give a good Gaussian curve. I said that 6 or 8 contributing errors (combined in each observation) give a good Gaussian curve.

Inst.-Capt. T. Y. Baker comments on the ambiguity of $\pm b$, when no indication is given as to whether b is mean square error, probable error, etc. One's knowledge of human nature is usually a sufficient guide—bearing in mind that the probable error is the smallest of the lot.

In reply to Mr Guild: The argument (both in the deduction of the Gaussian law for x, and elsewhere) assumes the errors to be independent. When errors occur in runs of one sign, as they sometimes do, I am afraid the calculator cannot do much—except blame the experimenter. In his criticism of the theory of errors, Mr Guild seems to have in mind mainly quantities obtained by direct measurement, where (after reading an account of the methods and precautions) one can judge from experience the sort of accuracy likely to be attained. But this is not typical of the method of least squares. I do not think anyone could form an idea of the accuracy of a quantity like h, which is indirectly determined, without going through the calculations. When a probable error has been calculated, we may employ our judgment of men and methods as to whether it should be accepted at its face value; but our judgment must be given something to work on.

In reply to Mr T. Smith: The admissible risk varies enormously with the instrument. In testing clinical thermometers you must consider what proportion of users you are prepared to kill by misleading the doctor. In testing taximeters you may justifiably take a somewhat higher risk of permitting overcharge. I am afraid the second question is too technical for me.

621.314.234:621.43.047

THE CONTROL OF IGNITION-COIL DISCHARGE CHARACTERISTICS

By G. I. FINCH and R. W. SUTTON
Imperial College of Science and Technology

Received November 2, 1932. *Read, with demonstration, January* 20, 1933.

ABSTRACT. The theory of the ignition coil is developed and is verified by means of the cathode-ray oscillograph. It is known that the secondary discharge consists of two parts, the "capacity" component due to the discharge, upon breakdown of the spark gap, of the energy stored in the secondary capacity, and the subsequent discharge of the electromagnetically stored energy, the so-called "inductance" component. It is shown that the spark-gap voltage is constant throughout the life of the inductance component, and that the inductance-component current is oscillatory but unidirectional.

The theory leads to the conclusion, confirmed experimentally, that the closing of the primary circuit during the life of the inductance component rapidly extinguishes the secondary discharge. A contact-breaker enabling the duration of opening of the primary circuit to be controlled with accuracy over a range between 0 and 4 msec. is described. The elimination of various portions of an ignition-coil discharge without material effect upon the remainder of the discharge is discussed.

§ 1. INTRODUCTION

It is well known that the ignition-coil discharge is a complex phenomenon, consisting in general of at least two parts, namely the so-called "capacity" component due to the discharge of the secondary self-capacity on the breakdown of the gap, and the subsequent discharge of electromagnetic energy stored in the secondary and known generally as the "inductance" component of the coil discharge. It is clear, therefore, that a systematic study of coil ignition would require that the rôles played by the capacity and inductive components be investigated individually. In order to do this, it would be necessary to be able to arrange conditions in such a manner that the discharge component not under examination could be completely suppressed and the other modified as required and in a known manner.

In what follows, an account is given, inter alia, of a method developed from the analysis of the ignition-coil circuit, by means of which the inductive component of the coil discharge can be either wholly or in part suppressed without modifying the preceding part of the discharge.

§ 2. THE THEORY OF THE IGNITION COIL WITH SPECIAL REFERENCE TO THE INDUCTIVE COMPONENT OF THE DISCHARGE

For analytical purposes, the circuit to be considered is that shown in figure 1. The conditions to be investigated are those which obtain immediately after the opening of the break.

The secondary-circuit resistance being neglected, the general circuit equations are as follows:—

$$L_1 di_1/dt + L_{12} di_2/dt + R_1 i_1 + e_1 = 0 \qquad \ldots\ldots(1a),$$

and

$$L_2 di_2/dt + L_{21} di_2/dt + e_2 = 0 \qquad \ldots\ldots(1b),$$

where L_1, L_2 and C_1, C_2 are the primary and secondary inductances and capacities respectively, R_1 is the primary resistance, i_1 and i_2 are the primary and secondary currents, e_1 and e_2 are the potential differences between the plates of the condensers C_1 and C_2, L_{12} and L_{21} are the coefficients of induction of the secondary on the primary and of the primary on the secondary respectively, and t is the time.

L_1, L_2, C_1, C_2
R_1, i_1, i_2
e_1, e_2
L_{12}, L_{21}
t

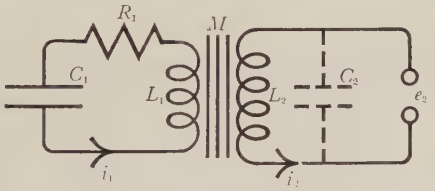


Figure 1.

The complete analysis of these equations has been carried out by Taylor-Jones* for two cases, namely (i) open-circuit secondary with no discharge passing and (ii) secondary shunted by an ohmic resistance, with the main object of determining the adjustments necessary in order to obtain the greatest possible secondary potential. The requirements governing the design of an ignition coil are, however, of a different nature, because such a coil is required to furnish a spark across a fixed gap the breakdown potential of which is considerably smaller than the maximum potential which the coil can deliver. Hence the solutions to the above general circuit equations, as given by Taylor-Jones, cannot be directly applied to the case of an ignition coil, where the secondary is shunted by a spark. A solution for this case could, however, be obtained, provided the (voltage, current) characteristic of the discharge were known.

The relationship between voltage and current in an induction-coil spark discharge does not appear to have been determined hitherto, although it is well known that the discharge cannot be even approximately represented by an ohmic resistance. Accordingly it was decided to determine the (voltage, current) characteristic of the discharge directly by means of a cathode-ray oscillograph.

Numerous oscillograms were taken of the secondary current and of the voltage across the electrodes of a gap traversed by the ignition-coil discharge, a wide range

* *The Theory of the Induction Coil* (London, Pitman and Sons, 1932).

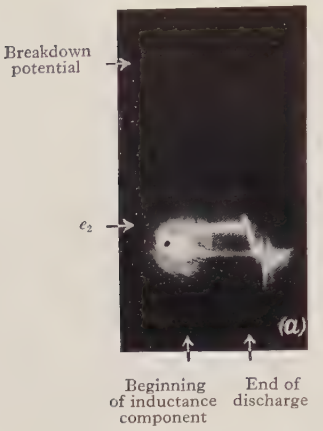
of gap-widths and primary currents being employed for this purpose. In obtaining the oscillograms of the secondary current, the oscillograph plates were connected across a 3040-Ω. non-inductively wound wire resistance in series with the spark gap and secondary. A suitably tapped 4-MΩ. Siemens Draht resistance across the spark gap served as potential-divider in securing the voltage oscillograms. Spherical (diameter 0·75 cm. brass), crossed-wire (no. 18-s.w.g. nickel or copper), and needle-point-sphere (needle-point the cathode and sphere the anode) spark gaps were employed.

A result typical of the range of conditions thus examined is reproduced in the oscillogram recorded in figure 4. This shows that the inductive component of the secondary current executes damped oscillations of considerable amplitude, but is throughout unidirectional. The voltage across the gap (figure 2 *a*) rises rapidly to a high value, the breakdown potential of the gap, and then falls with such extreme rapidity as to leave no trace on the oscillogram to a much lower value e_2, at which it remains practically constant throughout the remainder of the discharge. The oscillations occurring at the end of the voltage trace are not set up until the secondary current ceases to flow and are merely oscillations forced in the secondary. Thus the virtual constancy of the secondary voltage persists throughout the period during which the inductive component of the current is flowing through the gap, in spite of the fact that the current is varying over a wide range during this period. No sign of the secondary-current oscillations can be seen in the voltage trace, except in the extreme case when the gap-width is so large, or the primary current so small, that each single oscillation of the secondary current is followed by the extinction of the discharge. In this case, as is clearly seen in figure 2 *b*, the voltage rises afresh to the breakdown value of the gap at the beginning of each oscillation of the inductive component current, whereupon it immediately falls to the low constant value and remains there during the rest of the current oscillation.

The above oscillograms show clearly that as long as current is flowing the voltage across the gap is constant to within the practical limits of accuracy of the cathode-ray oscillograph employed.

i_0 On breaking the primary circuit, the primary current, which has an initial value of i_0, commences to fall, and the induced secondary current charges up the secondary capacity to the breakdown voltage of the gap. The secondary capacity thereupon discharges through the gap, forming the capacity component of the discharge. Its duration is so brief that, as is shown in figures 2 *a* and 2 *b*, the cathode-ray oscillograph fails to record a visible trace of its passage, although the fact that the inductive component current trace in figures 4, 6 *a* and 6 *b* never begins at, but always some distance above, the origin affords sufficient testimony to its existence. We are, therefore, justified in assuming that, for all practical purposes, from the moment of breakdown of the gap (when the primary

i_p current has a value i_p, which is less than the initial value i_0 by an amount depending upon the energy absorbed in charging up C_2 to the breakdown potential of the gap) the secondary voltage has a value e_2 which remains constant so long as the secondary current is flowing through the discharge and has a finite positive value.



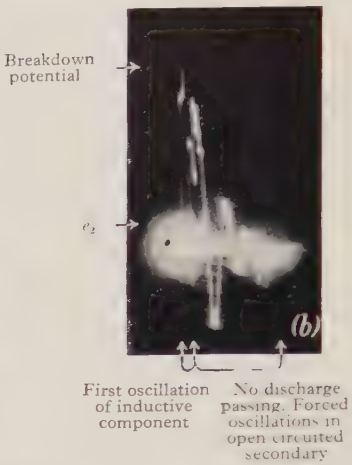


Figure 2.

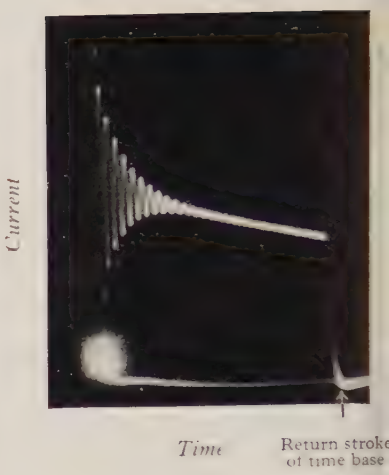


Figure 3.

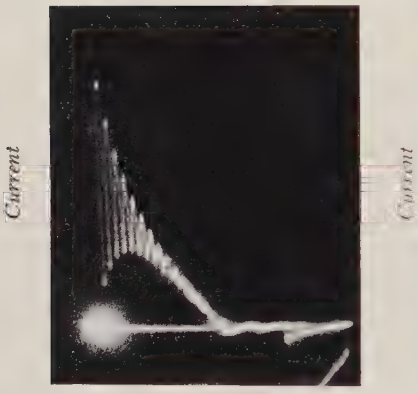


Figure 4.

Figure 5.

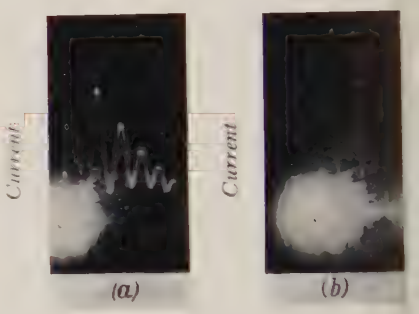
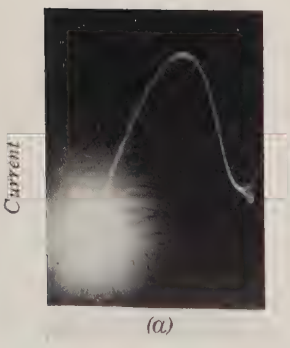
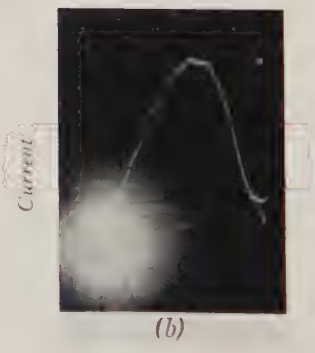


Figure 6.





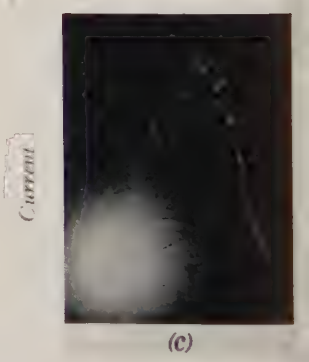


Figure 7.

To face page 291

Therefore no current flows in the secondary self-capacity throughout this period. It follows that the current-distribution in the secondary winding is uniform, and hence $L_{12} = L_{21} = M$. The general circuit equations can now be readily solved as follows:— M

$$L_1 d^2q_1/dt^2 + Md^2q_2/dt^2 + R_1 dq_1/dt + q_1/C_1 = 0,$$

and

$$L_2 d^2q_2/dt^2 + Md^2q_1/dt^2 + e_2 = 0,$$

where $q_1 = \int i_1 dt$ and $q_2 = \int i_2 dt$. q_1, q_2

These equations may be written in the form

$$(L_1 D^2 + R_1 D + 1/C_1)\, q_1 + MD^2 q_2 = 0 \qquad \ldots\ldots(2a),$$

and

$$MD^2 q_1 + L_2 D^2 q_2 = -e_2 \qquad \ldots\ldots(2b),$$

where D is the operator d/dt. D

Eliminating q_1, and putting $M^2 = \kappa^2 L_1 L_2$, we obtain κ

$$\{D^2 (D^2 + aD + b)/(1 + R_1 C_1 D + L_1 C_1 D^2)\}\, q_2 = -e_2/L_1 L_2 C_1 (1 - \kappa^2),$$

where

$$a = R_1/L_1 (1 - \kappa^2), \qquad a$$

and

$$b = 1/L_1 C_1 (1 - \kappa^2). \qquad b$$

The complete solution for q_2 is, therefore,

$$q_2 = A_2 + B_2 t - e_2 t^2/2L_2 + P_2 e^{m_1 t} + Q_2 e^{m_2 t},$$

where

$$m_1 = -a/2 + (a^2/4 - b)^{\frac{1}{2}}, \qquad m_1$$

$$m_2 = -a/2 - (a^2/4 - b)^{\frac{1}{2}}, \qquad m_2$$

and A_2, B_2, P_2, Q_2 are arbitrary constants. A_2, B_2, P_2, Q_2

Similarly, eliminating q_2 from equations $(2a)$ and $(2b)$ and substituting $M^2 = \kappa^2 L_1 L_2$, we obtain

$$(D^2 + aD + b)\, q_1 = Me_2/L_1 L_2 (1 - \kappa^2).$$

The complete solution for q_1 is, therefore,

$$q_1 = MC_1 e_2/L_2 + P_1 e^{m_1 t} + Q_1 e^{m_2 t},$$

where P_1 and Q_1 are arbitrary constants. P_1, Q_1

The solutions of the original differential equations are thus

$$q_1 = MC_1 e_2/L_2 + P_1 e^{m_1 t} + Q_1 e^{m_2 t} \qquad \ldots\ldots(3a),$$

and

$$q_2 = A_2 + B_2 t - e_2 t^2/2L_2 + P_2 e^{m_1 t} + Q_2 e^{m_2 t} \qquad \ldots\ldots(3b).$$

Differentiating these equations, we get

$$i_1 = m_1 P_1 e^{m_1 t} + m_2 Q_2 e^{m_2 t} \qquad \ldots\ldots(4a),$$

and

$$i_2 = B_2 - e_2 t/L_2 + m_1 P_2 e^{m_1 t} + m_2 Q_2 e^{m_2 t} \qquad \ldots\ldots(4b).$$

By substituting the expressions obtained for q_1 and q_2 in the two original differential equations, the following two relations between the six arbitrary constants are obtained:—

$$MP_1 + L_2 P_2 = 0 \qquad \ldots\ldots(5a),$$

and

$$MQ_1 + L_2 Q_2 = 0 \qquad \ldots\ldots(5b).$$

The remaining four relations follow from the initial conditions which are that, when the time $t = 0$,

q_p

$$q_1 = q_p = C_1 e_p,$$

$$q_2 = 0,$$

$$i_1 = i_p,$$

and

$$i_2 = i_s,$$

e_p where e_p is the potential difference between the plates of the primary condenser,

i_p, i_s and i_p and i_s are the primary and secondary currents, at the moment of breakdown.

The relations given by the initial conditions are

$$P_1 + Q_1 + MC_1 e_2/L_2 = C_1 e_p \quad \ldots\ldots(5c),$$

$$A_2 + P_2 + Q_2 = 0 \quad \ldots\ldots(5d),$$

$$m_1 P_1 + m_2 Q_1 = i_p \quad \ldots\ldots(5e),$$

and

$$B_2 + m_1 P_2 + m_2 Q_2 = i_s \quad \ldots\ldots(5f).$$

The six equations, $(5a)$ to $(5f)$, determine the arbitrary constants, as follows:—

$$A_2 = nC_1 (e_p - ne_2),$$

$$B_2 = i_s + ni_p,$$

$$P_1 = \{i_p + m_2 C_1 (ne_2 - e_p)\}/(m_1 - m_2),$$

$$Q_1 = \{i_p + m_1 C_1 (ne_2 - e_p)\}/(m_2 - m_1),$$

$$P_2 = -nP_1,$$

$$Q_2 = -nQ_1,$$

n where $n = M/L_2$.

In all practical cases, $b > a^2/4$, and $(a^2/4 - b)^{\frac{1}{2}}$ is, therefore, imaginary. On substituting the above values for the arbitrary constants and converting exponentials to circular functions, equations $(4a)$ and $(4b)$ become

$$i_1 = e^{-at/2} \{i_p \cos \omega t + I \sin \omega t\} \quad \ldots\ldots(6a),$$

and

$$i_2 = i_s + ni_p - e_2 t/L_2 - ne^{-at/2} \{i_p \cos \omega t + I \sin \omega t\} \quad \ldots\ldots(6b),$$

ω where $\omega = (b - a^2/4)^{\frac{1}{2}}$,

I and $I = \{2 (ne_2 - e_p)/R_1 - i_p\} a/2\omega.$

The voltage across the primary condenser is given by $e_1 = q_1/C_1$, i.e.

$$e_1 = ne_2 - e^{-at/2} \{(ne_2 - e_p) \cos \omega t - E \sin \omega t\} \quad \ldots\ldots(6c),$$

E where $E = i_p/\omega C_1 - (ne_2 - e_p)\, a/2\omega.$

These solutions can be written in the forms

$$i_1 = (i_p^2 + I^2)^{\frac{1}{2}} e^{-at/2} \cos (\omega t - \phi) \quad \ldots\ldots(7a),$$

$$i_2 = i_s + ni_p - e_2 t/L_2 - n (i_p^2 + I^2)^{\frac{1}{2}} e^{-at/2} \cos (\omega t - \phi) \quad \ldots\ldots(7b),$$

and

$$e_1 = ne_2 - \{(ne_2 - e_p)^2 + E^2\}^{\frac{1}{2}} e^{-at/2} \cos (\omega t + \theta) \quad \ldots\ldots(7c),$$

ϕ where $\tan \phi = I/i_p,$

θ and $\tan \theta = E/(ne_2 - e_p).$

The resistance of the secondary circuit has been neglected in deriving these equations which represent the inductive component of the discharge in terms of the coil constants, of the instantaneous values of primary and secondary current, and of primary voltage, at the moment of breakdown of the gap. The great simplification attained thereby is fully justified by the fact that the experimental results set forth below show, as was indeed to be expected, that the voltage-drop across the secondary-circuit resistance is always small in comparison with e_2.

Although exact numerical solutions of these equations cannot be obtained without a knowledge of the instantaneous values of the primary and secondary currents at the moment of breakdown of the spark gap, there are two particular cases for which they can be simplified. They are:—

Case 1. *Secondary short-circuited.* In this case $e_2 = 0$, $e_p = e_0$, and $i_p = i_0$, where e_0 and i_0 are the primary voltage and current respectively, at the moment of breaking the primary circuit. The equations now reduce to e_0, i_0

$$i_1 = (i_0^2 + I^2)^{\frac{1}{2}}\, e^{-at/2} \cos(\omega t - \phi),$$

$$i_2 = i_s + ni_0 - n\,(i_0^2 + I^2)^{\frac{1}{2}}\, e^{-at/2} \cos(\omega t - \phi),$$

and

$$e_1 = -\,(e_0^2 + E^2)^{\frac{1}{2}}\, e^{-at/2} \cos(\omega t + \phi).$$

Now,

$$I = -\,(i_0 + 2e_0/R_1)\, a/2\omega$$

$$= -\,i_0\,(1 + 2e_0/R_1 i_0)/\{4L_1(1 - \kappa^2)/R_1^2 C_1 - 1\}^{\frac{1}{2}},$$

and

$$E = i_0/\omega C_1 + ae_0/2\omega$$

$$= i_0/\omega C_1 + e_0/\{4L_1(1 - \kappa^2)/R_1^2 C_1 - 1\}^{\frac{1}{2}}.$$

Since, in practice, $R_1^2 C_1/4L_1(1 - \kappa^2)$ is less than 0·001, we may neglect I in comparison with i_0 and may write $E = i_0/\omega C_1$.

Moreover, since $e_0 \doteqdot i_0 R_1$,

$$e_0/E \doteqdot \omega R_1 C_1$$

$$\doteqdot [\{R_1^2 C_1/L_1(1 - \kappa^2)\}\,\{1 - R_1^2 C_1/L_1(1 - \kappa^2)\}]^{\frac{1}{2}}$$

$$\doteqdot 0.$$

Therefore e_0 may be neglected in comparison with E.

With the above approximations, the equations reduce to

$$i_1 = i_0 e^{-at/2} \cos \omega t \qquad \ldots\ldots(8a),$$

$$i_2 = ni_0\,\{1 - e^{-at/2} \cos \omega t\} \qquad \ldots\ldots(8b),$$

and

$$e_1 = (i_0/\omega C_1)\, e^{-at/2} \sin \omega t \qquad \ldots\ldots(8c).$$

It is thus apparent that in this case the primary current executes a series of damped oscillations about zero, having an initial amplitude equal to the current at break. The primary voltage executes damped oscillations of initial amplitude $i_0/\omega C_1$ about zero. The secondary current rises to a maximum value $2i_0 e^{-a\pi/2\omega}$ and executes damped oscillations of initial amplitude ni_0 about a mean value ni_0. In practice, an exponential decay curve must be superimposed on these oscillations, owing to the presence of resistance in the secondary circuit.

Case 2. Coil under working conditions. In this case the primary current is considerably in excess of the minimum value required to produce a secondary e.m.f. sufficient to break down the gap. The energy absorbed from the primary in charging up the secondary capacity to the breakdown potential of the gap is only a small fraction of the total primary energy. We may therefore assume that $i_s = 0$, $i_p = i_0$, and $e_p = e_0$. Since ne_2 is in general of the same order as e_0, we may, as in case 1, neglect I in comparison with i_0 and $(ne_2 - e_0)$ in comparison with E. The equations therefore become

$$i_1 = i_0 e^{-at/2} \cos \omega t \qquad \ldots\ldots(9a),$$

$$i_2 = ni_0 \{1 - e^{-at/2} \cos \omega t\} - e_2 t/L_2 \qquad \ldots\ldots(9b),$$

and

$$e_1 = ne_2 + (i_0/\omega C_1)\, e^{-at/2} \sin \omega t \qquad \ldots\ldots(9c).$$

The axis of the secondary-current oscillations is given by

$$i_2 = ni_0 - e_2 t/L_2 \qquad \ldots\ldots(10),$$

τ and the total duration τ of the secondary current by

$$\tau = Mi_0/e_2.$$

Thus the primary current behaves in exactly the same manner as in case 1. The secondary-current oscillations, however, occur about a linear axis given by equation (10). After the oscillations have been completely damped out, the secondary current continues to decay to zero along this line, the slope of which is given by $-e_2/L_2$. The total duration of the secondary discharge is Mi_0/e_2. It would appear, therefore, that with high primary currents and small gap-widths which give low values of e_2 the secondary discharge should continue for a considerable time after the oscillations have been damped out and the primary current has fallen to zero.

The above equations appear at first sight to indicate that i_2, after falling to zero, rises again in the reverse direction and eventually reaches an infinite value, while the primary voltage remains constant indefinitely at the value ne_2. It must be recalled, however, that the condition

$$e_2 = \text{constant}$$

applies only so long as i_2 has a positive value. Hence, as soon as i_2 falls to zero the equations no longer hold. At this point the primary voltage, starting at the value ne_2, executes a series of oscillations of small amplitude and of a frequency corresponding to that of the coil with an open secondary and at the same time induces oscillations of small amplitude in the circuit of the secondary and its associated self-capacity.

The above equations thus account satisfactorily for the form of the primary and secondary (current, time) traces obtained with the cathode-ray oscillograph. They show, in particular, that the secondary current, though oscillatory, flows in one direction throughout the life of the discharge and has a duration which is long compared with the natural period of the system. They also show that throughout the greater portion of the discharge the primary current is zero.

§ 3. APPARATUS

The following apparatus was employed:—

(i) An ignition coil, consisting of an m-shaped laminated stalloy core; a primary of 143 turns of no. 24-s.w.g. double-silk-covered copper wire wound in two layers on a cylindrical cardboard former 2·5 cm. in diameter; and a secondary of 10,000 turns of no. 24-s.w.g. enamelled copper wire wound on a former in 35 layers of 420 turns each and separated by waxed tissue paper. In the built-up coil, which could be readily taken apart, the primary with the secondary over it was placed over the middle limb of the core, the latter having a cross-section 1·5 cm. square. In what follows this coil will be referred to as "the coil," or as "the ignition coil."

(ii) A calibrated bank of condensers having mica dielectric and low-resistance conductors. The following capacities and any parallel combination thereof were available: 0·0577, 0·117, 0·237, 0·456, 0·90, and 1·79 μF.

(iii) A make-and-break commutator, with tungsten points, similar to that usually employed in internal-combustion-engine coil-ignition practice. This break, together with certain modifications carried out at a later stage, will be described in further detail below.

(iv) An Ardenne cathode-ray oscillograph and a quarter-plate camera fitted with a Ross Xpres f/1·9 lens.

(v) The time base previously described*.

(vi) A Cambridge triple-element oscillograph with one electrostatic and two electromagnetic elements.

§ 4. DETERMINATION OF THE COIL CONSTANTS

The constants of the built-up coil described above were measured by cathode-ray oscillographical methods similar in most respects to those described by Taylor-Jones† who, however, used an electrostatic oscillograph of special construction.

The chief constants required were the primary inductance and capacity, the secondary inductance and its distributed self-capacity, and the coefficient of coupling. The primary and secondary inductances of the assembled coil were measured on a Campbell inductometer, and also by means of a frequency-meter, at 2000 $\sim$. The coupling was obtained by the method employed by Taylor-Jones† from the relation $n_1^2/n_2^2 = (1 - \kappa^2)$ where n_1 and n_2 were the frequencies of the secondary with the primary circuit open and closed respectively. n_1, n_2

The following values for the principal coil constants were obtained:—

Primary inductance $L_1 =$ 0·00506 H.,

Secondary inductance $L_2 =$ 11·6 H.,

Coefficient of coupling $\kappa =$ 0·95.

* *Proc. Phys. Soc.* **43**, 502 (1931).

† *Loc. cit.*

§ 5. QUANTITATIVE VERIFICATION OF THE THEORY

Case 1. It was not practicable to obtain cathode-ray oscillograms of the secondary current when the secondary was short circuited. A 3040-Ω. resistance connected across the secondary sufficed, however, to provide the voltage necessary to operate the oscillograph satisfactorily.

An oscillogram typical of those obtained is shown in figure 3. The method of measurement has been previously outlined*, the accuracy attained being of the order of 2 per cent. In figure 3 the height of the mean line about which the initial oscillations occur is 23·1 mA., and the height of the first peak above this line is 19·6 mA. The logarithmic decrement of the damped oscillations is 0·16. Therefore, the superimposed exponential decay curve due to the secondary resistance being neglected, the equation to this curve is

$$i_2 = 23{\cdot}1 - 22{\cdot}7e^{-0{\cdot}16\,\omega t/\pi}\cos\omega t,$$

where i_2 is measured in milliamperes. In this case, the primary capacity and current were 0·237 μF. and 1·8 A. respectively. Similar results were obtained for different values of primary capacity, the current at break, however, being kept constant at 1·8 A. The results are summarized in table 1.

Table 1

C_1 (μF.)	A (mA.)	B (mA.)	λ
0·237	23·1	22·7	0·16
0·456	22·9	23·6	0·17
0·900	23·3	24·7	0·23
1·79	22·3	24·3	0·27

A, B The constants A and B refer to the general equation to the secondary current

$$i_2 = A - Be^{-\lambda\,\omega t/\pi}\cos\omega t,$$

λ and λ is the logarithmic decrement of the damped oscillations.

From the above results it may be concluded that the secondary current is independent of the primary capacity. According to the theory, the values of A and B should each be equal to mi_0, where $m = \kappa(L_1 L_2)^{\frac{1}{2}}$. On substituting the values of κ, L_1 and L_2, it is found that $m = 0{\cdot}02$ when $i_0 = 1{\cdot}8$ A. Hence $mi_0 = 36$ mA., a value which is considerably greater than the mean values of A and B, namely 23 mA. and 24 mA., actually measured in the case of the example under consideration. This, however, is not surprising, because the resistance of the secondary circuit was neglected in the theoretical treatment and in the practical case comprised not only that of the secondary winding itself, but also that of the 3040-Ω. shunt resistance which was required in order to obtain the potential necessary to operate the oscillograph.

It will be seen from table 1 that for small capacities the values of A and B are approximately equal, and that as the capacity is increased the value of A remains

* *Proc. Phys. Soc.* **44**, 190 (1932); **43**, 502 (1931).

approximately constant whilst the value of B increases steadily. This effect is evidently due to the fact that, owing to the presence of resistance in the secondary circuit, the damped oscillations occur about an exponential decay curve instead of about a horizontal straight line. As the capacity is increased the frequency of the oscillations likewise increases, and the first peak moves to a more steeply sloping part of the exponential decay curve. The value of B, which is determined by the height of the first peak, therefore tends to its theoretical value of 36 mA., but the value of A which is determined by the remainder of the oscillations remains constant.

Case 2. A series of (secondary current, time) oscillograms for various values of gap-widths was obtained. A typical result is shown in figure 4, from which it is clear that the axis of the oscillations, instead of being nearly horizontal as in case 1, is now a steeply sloping straight line (figure 4). For this reason the method* hitherto adopted for measuring the amplitudes of successive oscillations had to be somewhat modified, as follows:—

Let $a_1, a_2, a_3, \ldots$ be the amplitudes of successive oscillations, as measured from the axis. Let p be the constant vertical distance fallen by the axis in half a period. Then, if A_1 is the measured vertical distance between the first and second peaks, A_2 that between the second and third peaks, etc., we have

a_1, a_2, a_3 p A_1 A_2

$$A_1 = a_1 + a_2 + p,$$

$$A_2 = a_2 + a_3 - p,$$

$$A_3 = a_3 + a_4 + p, \text{ etc.},$$

or

$$A_1 = a_1(1 + l) + p,$$

$$A_2 = a_1 l(1 + l) - p, \text{ etc.},$$

where $1/l$ is the decrement, l

i.e. $$1/l = (A_1 - A_2)/(A_2 + A_3),$$

and $$a_1 = (A_1 + A_2)/(1 + l)^2.$$

The linearity of the axis of oscillation was well established by the close agreement between the values of the ratio l found after measuring up each oscillogram. As an example, it will suffice to give the results obtained from one oscillogram (no. 2, 3/6/31). The gap-width was 0·004 in., and the gap was between needle-points in air at atmospheric pressure. The primary current was 1·5 A., and the primary capacity 0·456 μF. The measured amplitudes were as shown in table 2.

The value obtained for a_1 is 23·9 mA., and the logarithmic decrement

$$\lambda = \log(1/l) = 0{\cdot}173.$$

The value for the initial amplitude, corrected for damping, is therefore 28·4 mA. The theoretical value for the initial amplitude is given by ni_0 and is 30 mA.

The oscillation-frequency was found by measurement to be 10,750 $\sim$, the value calculated from the theory being 10,500 $\sim$.

* See also § 9.

Table 2

m	A_m (mA.)	$A_m + A_{m+1}$ (mA.)	$1/l$
1	46	81	1·208
2	35	67	1·175
3	32	57	1·186
4	25	48	1·185
5	23	40·5	1·190
6	17·5	34	1·192
7	16·5	28·5	1·187
8	12	24	1·200
9	12	20	1·210
10	8	16·5	1·178
11	8·5	14	1·166
12	5·5	12	—
13	6·5	—	—
Mean value of $1/l$			1·189

The total duration of the discharge was found to be 1·23 msec. Now the total duration is given by the theory as

$$\tau = Mi_0/e_2 = nL_2i_0/e_2.$$

Therefore

$$e_2 = nL_2i_0/\tau.$$

Inserting the values $n = 0{\cdot}02$, $L_2 = 11{\cdot}6$ H., $i_0 = 1{\cdot}5$ A., we arrive at a theoretical value for e_2 of 283 V., as compared with a steady value of 295 V. found oscillographically.

Similar oscillograms were taken for various gap-widths, but for similar values of primary capacity and current at break. The results are given in table 3.

Table 3

Gap-width (in.)	Height of first peak (mm.)	τ (msec.)	e_2 calculated (V.)
0	56	—	—
0·004	51·5	1·2	290
0·008	53	1·1	316
0·015	56	0·86	405

It will be seen from the table that on opening the gap a sudden drop occurs in the height of the first peak of the secondary current. As the gap-width is further increased, however, the height of the first peak also increases. This observation is in accordance with the theory outlined above, since

$$i_2 = ni_0 - e_2t/L_2 - n\,(i_0^2 + I^2)^{\frac{1}{2}}\,e^{-at/2}\cos(\omega t - \phi).$$

The amplitude of the oscillatory part of i_2 therefore includes the term I where

$$I = \{2\,(ne_2 - e_0) - R_1i_0\}/2\omega L_1\,(1 - \kappa^2)$$

which increases with e_2.

It is also clear from table 3 that the total duration τ of the discharge decreases with increase in gap-width. This is in accordance with the expression,

$$\tau = ni_0 L_2/e_2,$$

derived above.

Case 3. We may further briefly consider the case in which the breakdown potential of the spark-gap is comparable with the maximum sparking potential afforded by the coil when supplied with a given primary current. In this case the theory outlined above is no longer valid.

Figure 6 *b* shows the (current, time) trace obtained with a primary current of 1·5 A., a primary capacity of 1·79 μF. and a gap-width of 0·008 in. The spark is now seen to consist of a series of separate discharges. In this particular example, the series is limited to three distinct discharges, all taking place in the same direction, and each being extinguished before the following one starts. The change from the discharge under the conditions of case 2 to this type of discharge was always found to take place suddenly whenever the primary current was reduced below a certain critical value depending upon the gap-width and frequency.

A feature of considerable interest, plainly visible in figure 6 *b* and also evident on the first half-oscillations of figure 4 and figure 6 *a*, is that a harmonic of small amplitude, low damping and high frequency is superimposed on the main oscillation in each case. Under the conditions of case 2, this harmonic is invariably present but is usually damped out before completion of the first half-cycle. Under the conditions of case 3 where the whole spark in effect consists of a series of discharges, one corresponding with each cycle, the harmonic is invariably prominent in the first half of each cycle and sometimes persists throughout the whole cycle, according to the conditions of primary current and gap-width. We have never been able to observe any trace of this harmonic when the secondary has been shunted by the 3040-Ω. resistance. The frequency of the harmonic is independent of changes in the primary circuit, but its amplitude increases with the gap-width. From the above observations we conclude that these oscillations are in some manner intimately associated with the nature of the spark discharge itself.

Figures 7 *a*, 7 *b* and 7 *c* show oscillograms in which the first (current, time) oscillations, obtained under the conditions of case 1, case 2 and case 3 respectively, are greatly extended in order to show the subsidiary oscillations more clearly. It will be seen that in case 1 no subsidiary oscillations occur and that in case 3 the wave-form of the subsidiary oscillations is distinctly saw-toothed.

§ 6. THE CAPACITY COMPONENT

The capacity component of the discharge has so far eluded direct observation by means of the cathode-ray oscillograph employed in this investigation, on account of its extremely high frequency and currents. Its existence and order of magnitude have, however, been determined indirectly. By making use of a short-wave receiver we found that, in the case of one particular arrangement of secondary leads and spark-gap, the most prominent frequency of the capacity-component oscillations

was of the order of 10^7 ~, though several disturbances of both higher and lower frequencies could be detected.

The oscillograms reproduced in figures 2 and 4 afford indirect evidence of the existence and nature of the capacity component. Thus in figures 2 *a* and 2 *b* the secondary-voltage trace is seen to rise rapidly to a maximum value corresponding to the breakdown potential of the gap, whereupon it falls with such extraordinary rapidity to the much lower constant value e_2, pertaining to the inductance component, that throughout this fall the trace is completely invisible. Further, it will be seen that in every case the (current, time) traces obtained under the conditions of case 2 are never seen to start from the origin. A gap invariably exists between the base line and the first visible part of the (current, time) trace. This gap, when observed visually, is seen to increase if the gap-width is increased.

From the general orders of magnitude of the capacities and inductances in the secondary circuit and the extremely high frequency of the capacity-component oscillations, it is evident that the oscillations are set up mainly in the circuit comprising the spark-gap S and the self-capacity C_2 of the secondary circuit including the leads. The initial amplitude of the current oscillations will therefore be given by VC_2p, where V is the breakdown potential of the gap, p is the pulsatance of the oscillations and C_2 is the self-capacity of the coil and leads. Assuming values of $10^{-4}\mu$F. for C_2, 15,000 V. for V and 12×10^6 ~ for the frequency, we obtain a value of 113 A. for the initial current amplitude, a value which is far in excess of the peak-current values of the order of between 10 and 100 mA. characteristic of the inductive component of the discharge. Because of this excessive amplitude and the high frequency of the oscillations, no record of this part of the discharge is seen on the oscillogram. Moreover, whilst the capacity component of the discharge is occurring, the inductive component of the current is steadily increasing from zero; and although it is swamped by the capacity component and so records no trace during that stage of the discharge yet, by the time when the capacity component has been damped out, the inductive component has grown to a small finite value. A gap is therefore seen between the first point in the trace and the base line. The length of this gap is proportional to V, the breakdown potential of the spark-gap, and therefore increases with the gap-width.

V, p
C_2

§ 7. CONTROL OF THE CAPACITY COMPONENT

The energy of the capacity component is determined solely by the self-capacity of the secondary circuit and the breakdown potential of the gap, i.e. the gap-width, and can therefore be easily varied over a wide range. Its frequency can also be controlled by the inclusion of further inductance in the secondary circuit. Practically complete elimination of the capacity component can be effected by including in the secondary circuit either a suitable resistance or, preferably, a diode run under such conditions as virtually to block the capacity component whilst passing the inductance component unimpaired.

A discharge in which the capacity component alone occurs, and from which the inductance component is completely absent, must necessarily be obtained when the primary current at break has the least value capable of producing a spark at the gap. A further method of completely eliminating the inductance component without affecting the capacity component is given below.

§ 8. CONTROL OF THE INDUCTANCE COMPONENT

The inductance component of the discharge has been shown to consist of a damped oscillatory current superimposed upon a linearly decaying current. The magnitude and duration of the inductance component are, as we have already seen, nearly directly proportional to the primary current at break and can, therefore, be readily controlled by adjusting the primary current. The inductance component can be completely eliminated, as indicated above, by adjusting the primary current to the least value necessary to produce a spark. The frequency of the oscillations can be controlled by varying the primary capacity. The logarithmic decrement, and with it the amplitude of the first inductance-component oscillation, can be varied by varying the resistance of the primary LC circuit.

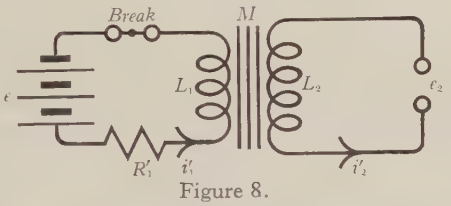


Figure 8.

It now remains to find a means of eliminating portions of the inductance component without affecting the remainder of the discharge. It might be desirable, for example, to examine the ignition properties of the first peak only of the discharge.

Suppose the primary circuit be first broken, thus producing a discharge in the secondary, and then closed again at some instant before cessation of the secondary discharge. The circuit conditions are then as shown in figure 8. The circuit equations are

$$L_1 di_1'/dt + M di_2'/dt + R_1' i_1' - e_0 = 0,$$

and

$$L_2 di_2'/dt + M di_1'/dt + e_2 = 0,$$

where i_1' and i_2' are the primary and secondary currents at any time after the closing of the primary circuit, and R_1' is the primary circuit resistance. These equations may be written in the form

$$(L_1 D + R_1')\, i_1' + M D i_2' = e_0 \qquad \ldots\ldots(10a),$$

and

$$M D i_1' + L_2 D i_2' = -e_2 \qquad \ldots\ldots(10b).$$

Eliminating i_1' and putting $M^2 = \kappa^2 L_1 L_2$, we obtain

$$[D\,(D + a')/\{1 - (M e_0 + L_1 e_2)\, D/R_1' e_2\}]\, i_2' = -a' e_2/L_2,$$

where $a' = R_1'/L_1\,(1 - \kappa^2)$.

The complete solution for i_2' is, therefore,

$$i_2' = A - e_2 t/L_2 + Be^{-a't},$$

A, B where A and B are arbitrary constants.

Similarly, eliminating i_2' from equations (10a) and (10b) and substituting M^2 for $\kappa^2 L_1 L_2$, we obtain

$$(D + a')\, i_1' = (e_0 + ne_2)\, a'/R_1'.$$

The complete solution for i_1' is, therefore,

$$i_1' = (e_0 + ne_2)/R_1' + Ce^{-a't},$$

C where C is an arbitrary constant.

The solutions to the original differential equations are thus

$$i_1' = (e_0 + ne_2)/R_1' + Ce^{-a't} \qquad \ldots\ldots(11a),$$

and

$$i_2' = A - e_2 t/L_2 + Be^{-a't} \qquad \ldots\ldots(11b).$$

By substituting these expressions in the original differential equations, the following relation between the three arbitrary constants is obtained:—

$$B + nC = 0 \qquad \ldots\ldots(12a).$$

The remaining two relations are obtained from the initial conditions which are that, when $t = 0$,

$$i_1' = i_1 \text{ and } i_2' = i_2,$$

i_1, i_2 where i_1 and i_2 are the instantaneous values of primary and secondary currents at the moment of remake. The relations given by the initial conditions are

$$C + (e_0 + ne_2)/R_1' = i_1 \qquad \ldots\ldots(12b),$$

and

$$A + B = i_2 \qquad \ldots\ldots(12c).$$

The three equations, (12a) to (12c), determine the arbitrary constants as follows:—

$$A = i_2 + ni_1 - n\,(e_0 + ne_2)/R_1',$$

$$B = n\,(e_0 + ne_2)/R_1' - ni_1,$$

and

$$C = i_1 - (e_0 + ne_2)/R_1'.$$

On substitution of the above values for the arbitrary constants, equations (11a) and (11b) become

$$i_1' = (1 - e^{-a't})\,(e_0 + ne_2)/R_1' + i_1 e^{-a't} \qquad \ldots\ldots(13a),$$

and

$$i_2' = (1 - e^{-a't})\,\{i_1 - (e_0 + ne_2)\}\, n/R_1' + i_2 - e_2 t/L_2 \qquad \ldots\ldots(13b).$$

If the primary circuit be closed when the oscillations have been damped out, or at a moment when the primary and secondary currents are passing through their mean values, then $i_1 = 0$, in which case

$$i_1' = (1 - e^{-a't})\,(e_0 + ne_2)/R_1' \qquad \ldots\ldots(14a),$$

and

$$i_2' = i_2 - e_2 t/L_2 - n\,(1 - e^{-a't})\,(e_0 + ne_2)/R_1' \qquad \ldots\ldots(14b).$$

Thus the secondary current falls rapidly to zero at the moment of re-establishment of the primary circuit. The term $e_2 t/L_2$, which represents the slow rate of fall of the secondary current obtained when the primary circuit is left open, is negligible

in all practical cases in comparison with the exponential term, and we can write, therefore,

$$i_1' = (i_0 + ne_2/R_1')\,(1 - e^{-a't}) \qquad \text{.....(15}a\text{)},$$

and

$$i_2' = i_2 - n\,(i_0 + ne_2/R_1')\,(1 - e^{-a't}) \qquad \text{.....(15}b\text{)},$$

because $i_0 = e_0/R_1'$.

Since i_2 can never be greater than ni_0, the term $n\,(i_0 + ne_2/R_1')$ is always greater than i_2. The secondary current therefore reaches zero after a finite time. Suppose that after interruption the primary circuit be re-established at a time equal to half the normal duration of the discharge, and that the oscillations have either been damped out before this point is reached or the primary and secondary currents are passing through mean values. Then,

$$i_2 = ni_0/2 \quad \text{and} \quad i_1 = 0.$$

Therefore

$$i_2' = ni_0/2 - n\,(i_0 + ne_2/R_1')\,(1 - e^{-a't}).$$

When $i_2' = 0$,

$$e^{a't} = (i_0 + ne_2/R_1')/(i_0/2 + ne_2/R_1').$$

Assuming the following values,

$$n = 0{\cdot}02,$$
$$i_0 = 2 \text{ A.},$$
$$e_2 = 360 \text{ V.},$$
$$R_1' = 3\ \Omega.,$$
$$L = 0{\cdot}005 \text{ H.},$$

and

$$\kappa = 0{\cdot}95,$$

then the value deduced for t is 48μsec. When the primary circuit is left open throughout the life of the discharge, then the time taken for the secondary current to fall to zero is given by $t = ni_0L_2/2e_2$. Assuming the value of 11·6 H. for L_2, the value obtained for t is 650μsec., a value which is 14 times that in the previous case where the primary circuit was closed during the life of the secondary discharge.

When it is remembered that, in calculating the duration of the secondary current after re-establishment of the primary circuit, the term e_2t/L_2 was neglected, and also that the decay of the secondary current occurs along an exponential curve, it becomes evident that an extremely rapid interruption of the secondary current should be obtainable.

A similar effect may be predicted to occur if remake, i.e. closing of the primary circuit, takes place at any other point along the secondary-current curve. If, for instance, remake be effected at the peak of an oscillation in the inductance component, then values of i_1 and i_2 can be readily calculated and the equation for i_2' found in the manner indicated above. In every case an equation is obtained which indicates that on remake the secondary current decays rapidly to zero in a time which is short in comparison with the total normal life of the discharge.

The total duration of the secondary discharge, which depends upon the value of i_0, under normal working conditions seldom exceeds 2 msec. Thus, in order

to test the above conclusions, it was necessary to have some form of contact-breaker which could be made to close after break in such a manner that the time during which the primary circuit was open could be continuously varied with sufficient precision over the range between 0 and, say, 4 msec. A first attempt to obtain this effect, by varying the maximum opening of the breaker contacts, failed as a result of irregular arcing due to the extremely small opening necessary. A second attempt, by recutting the contact-breaker cam surfaces, likewise proved unavailing because, owing to the exceedingly steep cam-surface angles required, it was not found possible to make the contact arm follow the cam throughout its travel.

Finally, the type of contact-breaker* described below and shown diagrammatically in figure 9 was evolved with satisfactory results.

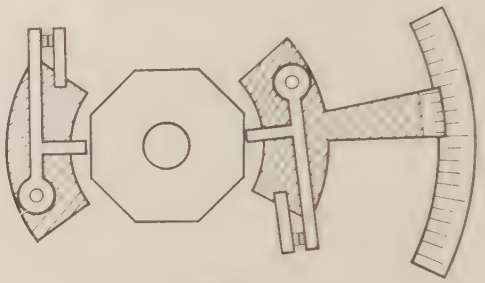


Figure 9.

The two pairs of contacts are shown on opposite sides of the cam. In practice they are arranged one above the other on the same side of the cam.

Two pairs of contacts in parallel were connected in the primary circuit. One pair of contacts was fixed in relation to the axis of an octagonal cam, about which the second pair could be rotated and secured in any required position. The angle of displacement between the pairs of contacts could be measured by means of an arc and vernier.

The operation of this contact-breaker was as follows. On rotation of the cam, with the moveable contact pair set out of phase with the fixed contact pair, one pair of contacts opened, but the other pair remained closed. The primary circuit was therefore still complete. On further rotation of the cam, the first pair of contacts began to close, but shortly before completion of the closure the second and hitherto closed pair opened, thus breaking the primary circuit which was afterwards remade by the closing of the first contact pair. This device made it possible to obtain any desired time of opening within the limits stated above and proved further to be free from sparking.

The results of some experiments carried out with the new contact-breaker are shown in figures 10 *a* to 10 *c*, in which are seen the effect of various times of opening upon the secondary current. It will be noticed particularly that on re-establishing the primary circuit during the life of the discharge the secondary current drops with great rapidity to zero, as predicted by the theory outlined above. Simul-

* B.P. 381,917 (1932).

taneously the primary current, as shown by the electromagnetic oscillogram, figure 11, rises rapidly to a definite value and then continues to rise, but at a much slower rate the time constant of which is characteristic of the ordinary type of remake of the primary occurring at any time after the cessation of the secondary discharge. That this is completely in accordance with the theory developed above will be seen if the primary current equations for early remake and for build-up under ordinary conditions are compared. The two equations are, respectively,

$$i_1' = (i_0 + ne_2/R_1')\{1 - e^{-R_1't/2L_1(1-\kappa^2)}\},$$

and

$$i_1' = i_0\{1 - e^{-R_1't/2L_1}\}.$$

It will be seen that the rate at which the primary current builds up has been greatly increased. The time constant of the circuit has been reduced from L_1/R_1' to $L_1(1 - \kappa^2)/R_1'$, i.e. to about a tenth of its normal value, and the primary current builds up as if approaching a final value of $(i_0 + ne_2/R_1')$ instead of i_0. The former value is clearly never reached in practice because the equations cease to hold as soon as the secondary current has fallen to zero which it does before the primary current has reached the value i_0. Upon complete extinction of the secondary discharge the primary current continues to build up towards its ultimate final value i_0, but at a rate now corresponding to a time constant L_1/R_1'.

Further secondary-current oscillograms were taken under conditions in which the primary circuit was remade at points intermediate between those shown in figures 10 *a* to 10 *e*. In all essentials similar results were obtained, and this shows that remake can be effected at any point on the secondary-current curve, either before, upon or after attainment of the secondary-current peak values, without affecting the essential features outlined above. Thus the value to which the primary current builds up at the rapid rate is independent of the instantaneous value at the moment of remake of the oscillatory component of the secondary current and depends only upon the value of the mean secondary current at the moment of remake

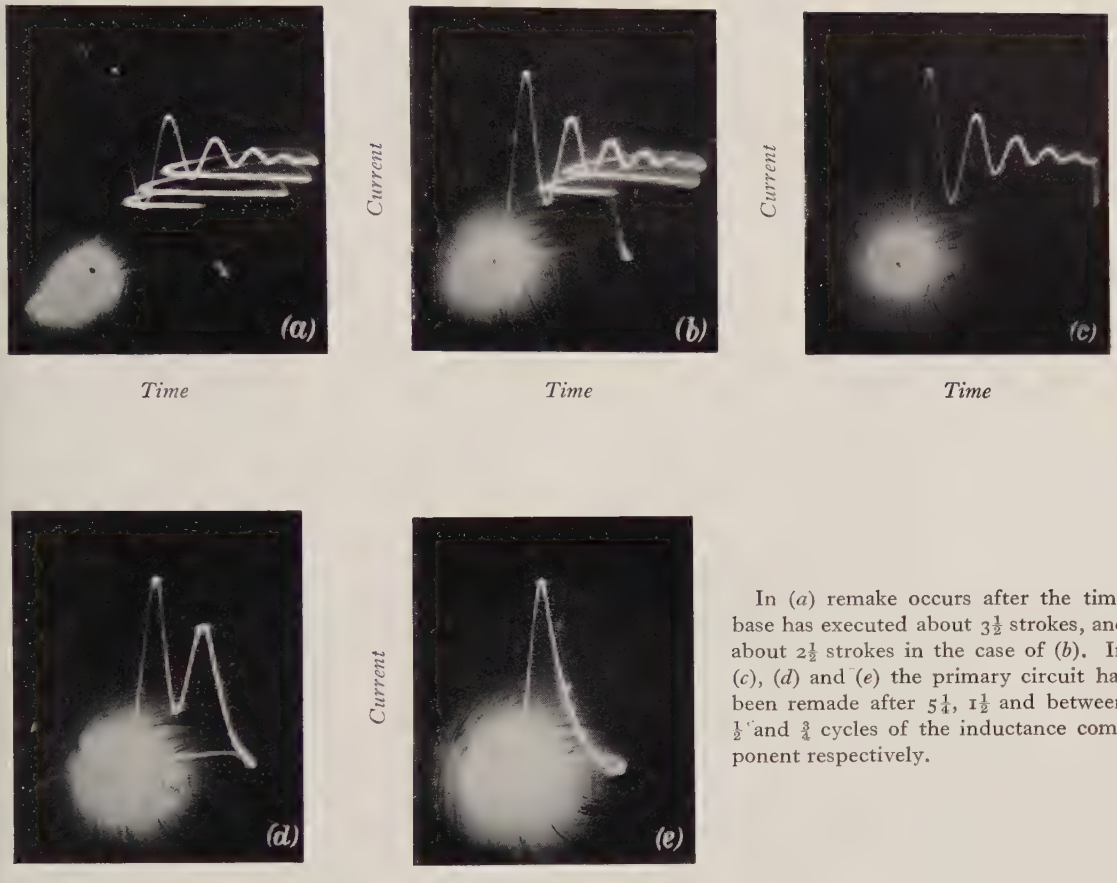
Thus complete control of the secondary discharge characteristics is possible. Those portions of the discharge whose igniting properties require examination can now be isolated. For example, the capacity component may be isolated with great precision by re-establishing the primary circuit at the beginning of the inductance component. Should it be desired to produce a discharge consisting of a single impulse, such as the first half-cycle of the inductance component, the break is set in such a manner that remake occurs at some instant after the first peak, but before the first trough in the inductance component oscillation, the capacity component being absorbed by means of a suitable resistance or diode connected in the secondary circuit.

§ 9. NOTE ON THE OSCILLOGRAM SHOWN IN FIGURE 5

In this oscillogram the (current, time) trace has been reversed by reversing the polarity of one pair of the oscillograph deflecting plates half-way through the exposure. By this means increased accuracy in the measurement of the oscillogram is obtained.

§ 10. ACKNOWLEDGMENTS

The above investigation was undertaken at the suggestion of the late Dr S. Z. de Ferranti, F.R.S., who was also actively associated with its prosecution until shortly before his death. We wish to place on record our deep sense of appreciation of Dr Ferranti's inspiring influence and of the liberal financial assistance afforded by Messrs Ferranti Ltd., without which the work could not have been carried out.



In (*a*) remake occurs after the time base has executed about $3\frac{1}{2}$ strokes, and about $2\frac{1}{2}$ strokes in the case of (*b*). In (*c*), (*d*) and (*e*) the primary circuit has been remade after $5\frac{1}{4}$, $1\frac{1}{2}$ and between $\frac{1}{2}$ and $\frac{3}{4}$ cycles of the inductance component respectively.

Figure 10.

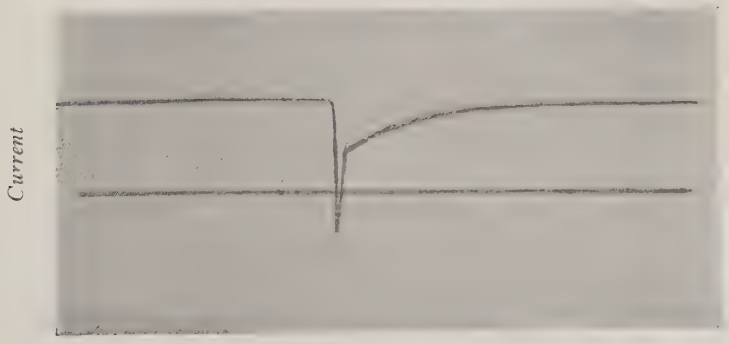


Figure 11.

To face page 306

551.585.1

THE WET-AND-DRY-BULB HYGROMETER: THE RELATION TO THEORY OF THE EXPERIMENTAL RESEARCHES OF AWBERY AND GRIFFITHS

By F. J. W. WHIPPLE, M.A., Sc.D., F.Inst.P.

Received November 23, 1932. *Read January* 20, 1933.

ABSTRACT. Simultaneous values have been published by Awbery and Griffiths of the dew-point, the vapour content as determined by a gravimetric method, and the reading of a wet-bulb thermometer, the observations covering a range of air-temperature from 30° to 100° C.

These observations have been rediscussed and it is demonstrated that August's formula by which vapour-pressure is deduced from dry- and wet-bulb readings is valid over the experimental range of temperature.

§ 1. INTRODUCTION

The wet-and-dry-bulb hygrometer is primarily apparatus for finding the humidity of the air, but it may be regarded also as apparatus for conducting a physical experiment demonstrating the cooling effect of evaporation. The recent paper* by Messrs Awbery and Griffiths, a study of observations of dry- and wet-bulb thermometers immersed in a stream of air and water vapour at temperatures ranging up to 100° C., meets adequately the need for tables by which humidities can be deduced from thermometer readings. The authors say nothing of the theoretical aspect of their observations, but by setting these out in detail they have given an opportunity, to anyone interested, to test such theories as may be suggested.

My examination of the observations leads to the conclusion that, within the order of accuracy of the investigation, the formula originally proposed by August and in general use for computing humidity at ordinary atmospheric temperatures is valid at high temperatures. Short tables of humidities computed by this formula and by a slightly modified formula have been prepared. These new tables may be compared with the final table of the paper by Awbery and Griffiths. As was to be anticipated, the differences between the tables are of no practical importance.

A notable result of the investigation is that at high temperatures very good ventilation of the wet bulb is not essential.

§ 2. AUGUST'S HYPOTHESIS AND THE RESULTING FORMULAE

The theory of the wet-bulb thermometer depends in the first instance on the conditions for equilibrium of the temperature of the thermometer bulb itself. The bulb is cooler than its surroundings. It receives heat from the air but it is also

* *Phys. Soc. Proc.* **44**, 115 (1932).

receiving heat by radiation from objects in the neighbourhood, by conduction down the thermometer stem and by the supply of fresh water. To take account of all these factors is practically impossible. All but the first are, however, comparatively unimportant when the air current passing the wet bulb is strong. Under such circumstances the wet bulb may be regarded as merely receiving heat from the air and giving up that heat as the latent heat of evaporation.

August's hypothesis* with regard to this exchange is that the air which gives up its heat to the wet bulb and thereby promotes evaporation is identical with the air which receives the evaporated moisture; the air becomes saturated at the wet-bulb temperature. The formula which will represent this hypothesis requires careful consideration if it is to be applied in extreme conditions with vapour pressure comparable with the total atmospheric pressure.

As a preliminary some discussion of latent heat is necessary. The latent heat of evaporation is naturally defined as the heat required to convert so much water into saturated vapour at the same temperature and under the pressure of that vapour alone. It is therefore important to demonstrate that the heat required is the same when the pressure is due to a mixture of air and water vapour.

V, P m, v L, E_1 E_2 Let the volume V be occupied by air and vapour at the pressure P. When the additional quantity m of water evaporates the volume is increased to $V + v$. Let mL be the latent heat under these circumstances. Further let mE_1 be the internal energy of the water to be evaporated and mE_2 the energy of the corresponding vapour. If E was the internal energy of the air and the original vapour, then the final value of the internal energy is $E - mE_2$. Further, the work done by the J expanding mixture is, in heat units, Pv/J, J being the mechanical equivalent of heat. Accordingly

$$mL = [E + mE_2 + Pv/J] - [E + mE_1],$$

so that

$$mL = m(E_2 - E_1) + Pv/J.$$

P_1 Now if P_1 is, in the final state, the partial pressure of the original air and water vapour, then by Boyle's Law

$$P_1(V + v) = PV.$$

P_2 Further, if P_2 is the partial pressure of the new vapour, $P_1 + P_2 = P$ and hence

$$Pv = P_2(V + v) = R'\theta' m,$$

R', θ' where R' is the gas constant for water vapour and θ' the temperature at which the process takes place. It follows that

$$L = E_2 - E_1 + R'\theta'/J.$$

Since P does not occur in this equation the value of L is not affected by the presence of the air. I am indebted to Mr D. Brunt for the enunciation and proof of this rule.

* It is doubtful whether August contemplated a continuous exchange. He refers to what has happened in the thin layer around the wet bulb. In this layer the air has evidently yielded up its heat to assist in the formation of water vapour. J. D. Everett, *Q.J.R. Met. Soc.* **18**, 13 (1892).

We proceed with the discussion of August's hypothesis. Let

P be the atmospheric pressure;

θ, θ' the dry- and wet-bulb temperatures; θ, θ'

e the vapour pressure of the atmosphere; e

e' the vapour pressure of saturated vapour at the temperature θ'; e'

c the specific heat of dry air at constant pressure; c

c' the specific heat of water vapour at constant pressure; c'

L the latent heat of evaporation at temperature θ'; L

n the ratio of the densities of water vapour and dry air at the same pressure and temperature. n

The hypothesis under consideration is that a certain sample of air, initially at temperature θ and pressure P and containing sufficient water vapour to exert a pressure e, is diluted with more water vapour; the mixture is at temperature θ' and pressure P, and contains sufficient water vapour to exert a pressure e'. The loss of heat is exactly balanced by the latent heat of evaporation of the additional vapour. Let M be the mass of dry air in the sample. The mass of vapour was initially M
$Mne/(P-e)$. In the final state the mass of vapour associated with M is $Mne'/(P-e')$. The additional quantity of water vapour is therefore m, where

$$m = \left(\frac{e'}{P-e'} - \frac{e}{P-e}\right) Mn = \frac{P(e'-e)}{(P-e)(P-e')} Mn.$$

The exchange of heat is equivalent to that taking place in the following stages: (1) the original air and water vapour are brought to the temperature θ', the pressure remaining constant; (2) the additional water vapour is evaporated under the pressure P into the atmosphere at the temperature θ'.

By hypothesis the heat communicated to the wet bulb by radiation, etc. is to be ignored. Accordingly the equation of equilibrium is

$$\frac{(e'-e)\,PnL}{(P-e)(P-e')} = (\theta - \theta')\left[c + \frac{e}{P-e} nc'\right].$$

Since in c.g.s. units $c = 0{\cdot}24$, $c' = 0{\cdot}42$ and $n = 0{\cdot}622$, it follows that $nc'/c = 1{\cdot}09$. As long as e is considerably less than P, the difference between this number 1·09 and unity may be ignored without serious error. With that simplification the equation of equilibrium reduces to

$$e' - e = B_1 P(\theta - \theta'),$$

where
$$B_1 = \left(1 - \frac{e'}{P}\right)\frac{c}{nL}. \qquad B_1$$

In the usual and historic statement of the theory the factor $(1 - e'/P)$ does not appear, and the psychrometric equation is then

$$e' - e = B_0 P(\theta - \theta'),$$

where
$$B_0 = c/nL. \qquad B_0$$

An expression including a factor equivalent to $(1 - e'/P)$ is given* however by Apjohn, a contemporary of August's.

The distinction between B_0 and B_1 will become very important when the temperature of the wet bulb approaches the boiling-point. Numerical values of B_0 and B_1 are given below (p. 314).

§ 3. TAYLOR'S THEORY

August's hypothesis, at any rate in the explicit form adopted here, is obviously artificial. In reality the air current cannot be divided sharply into two streams, one of which becomes saturated whilst the other is unaffected by the presence of the wet bulb. A more complicated theory, proposed by G. I. Taylor, was communicated by him to Principal Skinner, who published it in the article on humidity in the *Dictionary of Applied Physics*, III (1923). Taylor's hypothesis is that, whilst the air-flow is generally turbulent, there is a film of non-turbulent air passing slowly over the wet bulb. The diffusion of heat and of moisture are of different characters in the turbulent layer and in the non-turbulent film.

θ, e, θ', e' Let θ and e be the temperature and vapour-pressure in the free air; θ' and e'
θ'', e'' are the values at the surface of the wet bulb; θ'' and e'' are the values at the boundary
z, k between the film and the turbulent air. Let z be the thickness of the film, k the
D conductivity of air and D the diffusivity of water vapour through air. Following
ρ Skinner we regard the density of air as a constant ρ. The density of water vapour at pressure e is accordingly $n\rho e/P$.

The rate at which heat is conducted inwards to the wet bulb is per unit area

$$k(\theta'' - \theta')/z.$$

The rate at which water vapour diffuses outwards from the density $n\rho e'/P$ to the density $n\rho e''/P$ is

$$Dn\rho(e' - e'')/Pz.$$

As the incoming heat evaporates the water,

$$k(\theta'' - \theta') = Dn\rho(e' - e'')L/P.$$

E Now let E be the rate at which eddies are replacing the cool damp air of the outside of the film by the warmer, drier, natural air. E is reckoned as volume per unit area and unit time.

The rate at which heat is carried inwards is

$$E\rho c(\theta - \theta'').$$

The rate at which water vapour is carried outwards is

$$En\rho(e'' - e)/P.$$

The physical significance of E is rather vague; the essential assumption is that the coefficient E has the same value in both these statements.

Since the flow of heat inwards and of water vapour outwards are both continuous, it follows that

$$E\rho c(\theta - \theta'') = En\rho(e'' - e)L/P.$$

Thus

$$e'' - e = P(\theta - \theta')c/nL,$$

* *B.A. Report* 1843, p. 36.

whilst $$e' - e'' = \frac{P(\theta'' - \theta')}{nL} \frac{k}{\rho D}.$$

To eliminate e'' and θ'' simultaneously between these last two equations is only possible if

$$c = k/\rho D.$$

Now it so happens that $$k/\rho D = 0{\cdot}22,$$

whilst $$c = 0{\cdot}24.$$

The quoted values of k, ρ and D refer to the temperature 0° C., but it is known that $k/\rho D$ is nearly independent of temperature. We may ignore the difference between 0·22 and 0·24 and so obtain the equation

$$e' - e = B_0 P(\theta - \theta')$$

with $B_0 = c/nL$ as before.

According to Taylor dry-and-wet-bulb hygrometry is possible only because the diffusivity of water vapour through air and the conductivity of air are related in such a special way. Otherwise under varying conditions the changes in temperature and humidity in the eddy layer and in the non-turbulent layer would not be in proportion and there would not be a definite wet-bulb temperature corresponding with a given hygrometric state of the air.

It is possible to analyse more closely the details of Taylor's theory. For instance we can take account of the continual displacement of water from the outside of the non-turbulent film and regard θ'' and e'' as averages. We can investigate the diffusion of vapour through the film and allow for the fact that the gradient is not uniform. The gradient of temperature is also not uniform. Further we might allow for the fact that the density of the air is not constant. The general result of all these refinements is that the psychrometric formula with the constant B_0 is not exact and that the formula involving B_1 is likely to be nearer to the truth.

Experiments with water on the wet bulb in the ordinary atmosphere will not enable us to discriminate between August's theory and Taylor's. By the use of other liquids and gases it has been demonstrated that Taylor's is the more satisfactory.

§ 4. THE EXPERIMENTS OF AWBERY AND GRIFFITHS

Awbery and Griffiths exposed their dry- and wet-bulb thermometers in a current of humidified air at atmospheric pressure. They determined the humidity in two ways, by the dew-point and by a gravimetric method. They give their results as relative humidities corresponding with given readings of the dry- and wet-bulb thermometers. For comparison with theory we require the vapour pressure. The method of reduction may be illustrated by the first entries in table 1 of the paper*.

* The vapour pressures are taken from the *Wärmetabellen* of Holborn, Scheel and Henning (Braunschweig, 1919).

Dry-bulb temperature	$= \theta = 100{\cdot}6^\circ$ C.
Wet-bulb temperature	$= \theta' = 75{\cdot}0^\circ$ C.
Relative humidity by dew-point method	$= r_1 = 35{\cdot}3\ \%$.
„ „ by gravimetric method	$= r_2 = 35{\cdot}6\ \%$.
Saturation pressure at temperature θ	$= E = 776{\cdot}4$ mm.
Vapour pressure by dew-point method	$= e_1 = r_1 E = 274$ mm.
„ „ by gravimetric method	$= e_2 = r_2 E = 276$ mm.
„ „ over wet bulb	$= e' = 289$ mm.

We require the coefficient B in the formula

$$e' - e = BP(\theta - \theta').$$

Assuming $P = 760$ mm. and substituting from the foregoing data, we find:

B_d By dew-point method $B_d = 8 \times 10^{-4}$.

B_g By gravimetric method $B_g = 7 \times 10^{-4}$.

The estimates of B_d and of B_g are plotted in figures 1 and 2 respectively, the co-ordinates being θ' and $(\theta - \theta')$. A few anomalous results have not been plotted. It will be seen that there is no very systematic distribution of the entries. B_d tends to be small when $(\theta - \theta')$ is small and θ' is high, but there is no similar tendency with B_g.

The estimates exceeding 9×10^{-4} and less than 2×10^{-4} happen to be equally numerous. When these are excluded the mean values of B_d and B_g are $6{\cdot}3 \times 10^{-4}$ and $6{\cdot}9 \times 10^{-4}$ respectively.

It follows that the observations as a whole can be represented by the formula

$$e' - e = BP(\theta - \theta')$$

with

$$B = 6{\cdot}6 \times 10^{-4},$$

but the question whether B is really constant must be left open.

It is remarkable that whilst the observations of Awbery and Griffiths yield such good curves when plotted in their paper with relative humidity and dry-bulb temperature as variables, the same observations give the irregular distribution of numbers in figures 1 and 2. The reason is that the wet-bulb temperature and the dew-point differ but slightly, so that a small error in either, an error which hardly affects the relative humidity, has a large influence on B, the variable of the present paper. In the example cited above if the distinction between wet-bulb temperature and dew-point had been ignored the relative humidity would have been computed from the wet-bulb reading as e'/E or 37 per cent. As the true value was 35 per cent the error would for most purposes have been negligible.

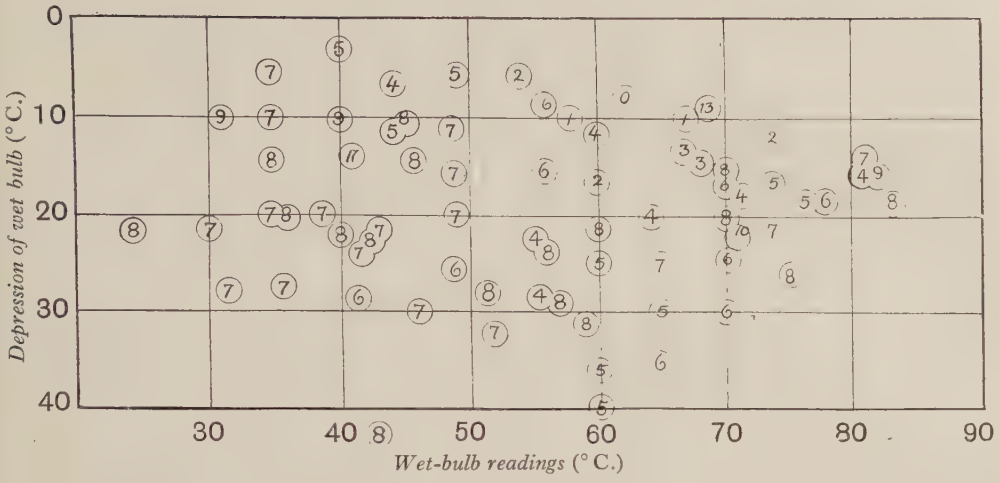


Figure 1. Values of $10^4 B$ as determined by dew-point observations.

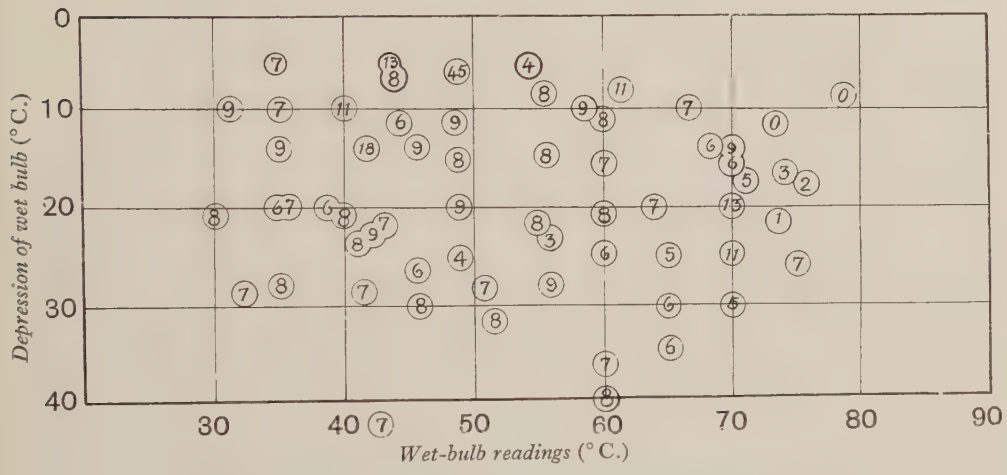


Figure 2. Values of $10^4 B$ as determined by absorption method.

§ 5. THEORY AND EXPERIMENT

To sum up the results, I have set out in table 1 the values of the coefficient B given by theory and by observation and also those adopted in certain tables.

The theoretical values are those given by the formulae for B_0 and B_1 in the present paper. The values adopted* for the latent heat are Henning's†. The best set of outdoor experiments for deriving the coefficient B is that made by Prof.

* The values of latent heat used by August, Apjohn and Regnault are quoted by Everett, *Q.J.R. Met. Soc.* **18** (1892). All used 0·267 for c.

† *Wärmetabellen*, p. 67, quoting from *Ann. d. Phys.* **29**, 441 (1909).

Marvin and other observers in the United States, some at considerable altitudes and therefore with low values of P, and discussed* by Prof. Ferrel. Marvin used sling thermometers and found the dew-point by Regnault's hygrometer. Ferrel derived from the observations the formula

$$B = 6{\cdot}56 \times 10^{-4}(1 + {\cdot}0019\theta').$$

This formula has been used to obtain the results against his name in table 1.

Table 1. The psychrometric constant, in theory and in practice

Temperature (° C.)	0	10	20	30	40	50	60	70	80	90	100
of wet bulb (° K.)	273	283	293	303	313	323	333	343	353	363	373
Vapour pressure (mm.)	4·6	9·2	17·5	31·8	55	93	149	234	355	526	760
(mb.)	6·1	12·3	23·4	42·5	74	123	199	312	473	701	1013
Latent heat (cal./gm.)	595	590	585	580	575	569	563	558	552	545	539
	Values of B (c.g.s. units $\times 10^{-4}$)										
Theory:											
B_0 or c/nL	6·49	6·54	6·60	6·66	6·72	6·79	6·86	6·92	7·00	7·08	7·16
B_1 or $(1 - e'/P)\,B_0$	6·45	6·46	6·45	6·38	6·23	5·97	5·52	4·79	3·73	2·19	0·00
Observation:											
Marvin-Ferrel	6·56	6·68	6·81	6·93	—	—	—	—	—	—	—
Svennson	←	6·45	→	—	—	—	—	—	—	—	—
Awbery and Griffiths:											
Dew-point	—	—	—	←	—	6·3	—	→	—	—	—
Gravimetric	—	—	—	←	—	6·9	→	—	—	—	—
Tables:											
Smithsonian (Ferrel)	6·60	6·68	6·75	6·83	—	—	—	—	—	—	—
Assmann (Sprung)	←	—	6·62	→	—	—	—	—	—	—	—
Accompanying tables 2, 3	—	—	—	←	—	6·6	—	—	—	—	→

Note. The theoretical values of B are found on the assumptions that $c = 0{\cdot}24$, $n = 0{\cdot}622$.

Accepting the theory that B ought to vary inversely as the latent heat, Ferrel found the alternative formula

$$B = 6{\cdot}60 \times 10^{-4}(1 + {\cdot}00115\theta').$$

This is the basis of the psychrometric tables† used in the United States.

Of laboratory experiments the best known are those of Svennson‡ who aspirated air past his thermometers in a vessel in which the air-pressure was sometimes reduced as low as 300 mm. The vapour pressure was determined by Sonden's apparatus in which the actual vapour pressure of a sample of the atmosphere can be measured directly. Svennson's value of B is lower than that adopted by most observers. He considers that this is because there are systematic errors in dew-point determinations and in the absorption method. These tend to give too low values for e. It will be noticed that Svennson's value of B is in good agreement with our formula for B_1.

* *Annual Report, Chief Signal Officer, Washington*, App. 24 (1886).
† *Smithsonian Meteorological Tables*, tables 82, 84 (5th Edition, 1931).
‡ *Met. Z.* **13**, 201 (1896).

The tables adopted* for the Assmann psychrometer are based on the simple formula given by Sprung†

$$e' - e = 0{\cdot}5\,(\theta - \theta')\,P/755.$$

The equivalent value of B is $6{\cdot}62 \times 10^{-4}$.

There has been some tendency in the past to lay undue stress on the differences between the values of B given by one investigator and another and to say that empirical tables are better than any formula for deriving humidity from wet-bulb readings. As will be seen from table 1, the range of values of the coefficient B for ordinary temperatures is quite small. There is, it is admitted, considerable uncertainty about the best use to be made of the readings of wet-bulb thermometers which are not "well ventilated" but that uncertainty need not affect confidence in the interpretation of the readings of such thermometers as *are* well ventilated.

It is satisfactory that the new determinations of B for moderate temperatures are in agreement with the older determinations for low temperatures. The new observations are not good enough, however, to discriminate between the merits of the formulae in which the coefficients are B_0 and B_1.

§ 6. NEW TABLES

To cover the range of observations with dry-bulb temperatures up to 100° C., skeleton tables of vapour pressure and relative humidity have been prepared.

Our theoretical discussion led to the conclusion that the coefficient B should be variable and a function of the wet-bulb temperature, whilst the observations were satisfied well enough by the assumption that B was constant. Tables based on each of these hypotheses have been prepared. It might have been anticipated that the tables computed on the two assumptions would be utterly inconsistent. As a matter of fact, the differences are insignificant. The reason is that at high temperatures the difference between the wet-bulb temperature and the dew-point is small, and the results of using alternative values of the factor B cannot be distinguished. Further, the agreement between the tables of relative humidity and that of Awbery and Griffiths is almost perfect. The only discrepancy is where they had extrapolated for wet-bulb depressions beyond the range of experiment.

It is a practical deduction from these comparisons that in psychrometric observations at high temperatures strong ventilation is not a necessity. This must be true, since the computed humidities would be affected but slightly if Regnault's moderate-wind value were used for B, instead of the strong-wind value. On the other hand I should like to express the hope that experiments will be made to fix the values of B more precisely. I would suggest for consideration the use of a differential thermometer to determine the interval between the dew-point and the wet-bulb temperature. If room air were used without the admixture of steam, the experiment would be comparatively easy to control, and consistent values of the psychrometer constant B, accurate to two significant figures, might be obtained.

* *Aspirations-Psychrometer-Tafeln* (Braunschweig, 1908). † *Das Wetter* (1888).

Table 2*. Vapour pressure derived from psychrometer readings by the use of the empirical formula, $B = 6{\cdot}6 \times 10^{-4}$

Dry-bulb temperature (° C.)	30	40	50	60	70	80	90	100
Wet-bulb depression (° C.)	Vapour pressure (mb.)							
0	42·5	73·8	123·3	199	312	473	701	1013
5	28·4	53·0	92·6	154	247	382	575	842
10	16·7	35·8	67·1	117	192	305	467	694
15	7·1	21·7	46·3	86	148	240	375	568
20	—	10·0	29·1	60	110	186	299	460
25	—	0·4	15·0	40	79	141	234	368
30	—	—	3·3	22	54	103	179	292
35	—	—	—	8	33	73	134	227
40	—	—	—	—	16	47	97	172
45	—	—	—	—	2	26	66	127

Table 3*. Relative humidity derived from psychrometer readings by the use of the empirical formula, $B = 6{\cdot}6 \times 10^{-4}$

Dry-bulb temperature (° C.)	30	40	50	60	70	80	90	100
Wet-bulb depression (° C.)	Relative humidity (per cent)							
0	100	100	100	100	100	100	100	100
5	67	72	75	77	79	81	82	83
10	39	49	54	59	62	65	67	69
15	17	29	38	43	47	51	54	56
20	—	14	24	30	35	39	43	45
25	—	1	12	20	25	30	33	36
30	—	—	3	11	17	22	26	29
35	—	—	—	4	11	15	19	22
40	—	—	—	—	5	10	14	17
45	—	—	—	—	1	5	9	13

* Tables 2 and 3 are computed for normal atmospheric pressure; $P = 1013$ mb.

Table 4*. Vapour pressure derived from psychrometer readings by the theoretical formula, $B_1 = (1 - e'/P)\, c/nL$

Dry-bulb temperature (° C.)	30	40	50	60	70	80	90	100
Wet-bulb depression (° C.)	Vapour pressure (mb.)							
0	42·5	73·8	123·3	199	312	473	701	1013
5	28·5	53·1	92·8	155	248	383	576	844
10	16·9	36·0	67·5	117	193	307	469	699
15	7·3	21·9	46·7	87	149	243	379	573
20	—	10·3	29·6	61	111	188	302	465
25	—	0·7	15·4	40	80	143	237	374
30	—	—	3·8	23	55	105	182	297
35	—	—	—	9	34	74	137	232
40	—	—	—	—	17	49	99	177
45	—	—	—	—	2	30	68	131

Table 5*. Relative humidity derived from the values of vapour pressure in table 4

Dry-bulb temperature (° C.)	30	40	50	60	70	80	90	100
Wet-bulb depression (° C.)	Relative humidity (per cent)							
0	100	100	100	100	100	100	100	100
5	67	72	75	78	80	81	82	83
10	40	49	55	59	62	65	67	69
15	17	30	38	44	48	51	54	57
20	—	14	24	31	36	40	43	46
25	—	1	12	20	26	30	34	37
30	—	—	3	12	18	22	26	29
35	—	—	—	5	11	16	20	23
40	—	—	—	—	5	10	14	17
45	—	—	—	—	1	6	10	13

* Tables 4 and 5 are computed for normal atmospheric pressure; $P = 1013$ mb.

DISCUSSION

Dr Ezer Griffiths. I am gratified to hear that Dr Whipple has been able to verify the theory of the wet-and-dry-bulb thermometer by means of our experimental observations. It is a very interesting point that the formula holds good over such a wide range of temperature. Has the author considered the case of dry-bulb temperatures below 0° C.? We are engaged at present in a study of hygrometers at low temperatures. In this investigation the wet bulb is covered with a sheath

of ice. We use various forms of the wet-and-dry-bulb hygrometer. In some the wet-and-dry bulbs take the form of a differential thermo-element with 33 wet and dry junctions. In others differential resistance thermometers are used. The wet-and-dry-bulb method is compared with the gravimetric and dew-point methods, both of which present considerable experimental difficulties at low temperatures.

Mr J. H. Awbery. I think Dr Griffiths agrees with me in that the wet-and-dry-bulb formula has never seemed to us to be based on a satisfactory theory. Even now, with the improved derivation given by the author of this paper, its basis includes the equality of 0·22 with 0·24; and we have no investigation of the errors which arise by ignoring the difference between these two quantities. Whatever may be the basis of the formula, its practical utility cannot really be denied. I had never compared direct observation with the formula until recently, but the present paper led me to analyse our own results in the light of the equation. I am now completely converted to the view that for smoothing observations it is perfectly satisfactory: but my adherence to orthodoxy is on account of empirical evidence, and is not due to an inward realization of truth.

That my conversion is real, may be inferred from the fact that the detailed table which Dr Griffiths and I have prepared * was constructed in part with the assistance of the formula.

Mr C. S. Wright. I should like to ask Dr Whipple whether meteorologists place any reliance on the readings of wet-and-dry-bulb thermometers in moderately low temperatures? The last Scott Expedition observed a condensation in the winter on ice surfaces, when the wind velocity was low. Kidston in his analysis of the results of Shackleton's first expedition explains the numerous occasions of a *higher* wet-bulb than dry-bulb temperature as due to a similar condensation on the ice surrounding the "wet" bulb. May not condensation also occur above freezing-point?

Author's reply. My answer to Dr Griffith's question is that I see no reason to doubt that a formula of August's type should be valid for the case in which the wet bulb is covered with ice. Meteorologists use different formulae for ice-covered and water-covered bulbs (allowing for the greater latent heat of evaporation from ice) and also recognize the distinction between the vapour pressure of air saturated with respect to ice and to water.

In Observatory practice we find that the wet-bulb thermometer is frequently below the freezing-point for many hours before the water on the muslin freezes. Moreover freezing takes a considerable time, and so does thawing. On account of uncertainties as to the state of the covering of the wet bulb, it is found better to trust to the hair hygrograph for finding humidity of air near the freezing-point.

For finding humidities at the low temperatures encountered in aeroplane ascents the Meteorological Office provides observers with warm water in a thermos flask. Between observations the wet bulb is dipped into the water and a new coating of ice is soon formed.

* *National Physical Laboratory, Annual Report for* 1932.

No satisfactory method has been evolved for determining the humidity at great heights—for example, in the stratosphere, where temperature is comparable with 220° K. Saturation pressure at that temperature is only 0·027 mb. as compared with 6·1 mb. at the freezing-point. I hope that Messrs Awbery and Griffiths will be able to make some practical suggestion. We want to know more about humidity in the stratosphere because we believe that the absorption of long-wave radiation by this water vapour plays a very important part in meteorology.

In reply to Mr C. S. Wright: I am rather sceptical about wet-bulb readings higher than the dry-bulb readings. I wonder whether there is any evidence that the thermometers used in Shackleton's expedition had been recalibrated before the paradoxical observations.

Condensation on ice surfaces in the open is a common enough phenomenon, the ice surface being cooled below air temperature. One can sometimes discriminate on a frosty morning between frozen dew and the crystalline ice deposited on the dew. I do not see how this explains the rise of wet-bulb temperature above dry-bulb temperature, both temperatures being taken, presumably, with screened thermometers.

The standard hygrometric tables* in use in Sweden are based on experiments by Ekholm, who attributed high wet-bulb readings to the hygroscopic qualities of muslin. According to these tables the equality of dry- and wet-bulb readings connotes a relative humidity of 97 per cent. I hope that Ekholm's work will be critically examined before the final results of the N.P.L. investigation are published.

One reader of the proof of my paper has raised the question whether the cooling of the water fed to the wet bulb should not be allowed for in the theory. One answer to this question is that the well-ventilated psychrometer in its best known form, that due to Assmann, has not a continuous feed, while in other cases the feed is along a wick which is exposed to the air so that the water is brought very nearly to wet-bulb temperature before reaching the muslin. The more fundamental answer is that even if the water reached the wet bulb at dry-bulb temperature and gave up heat as it spread over the muslin, the allowance for this heat would not seriously affect our analysis. The amendment would be equivalent to a reduction in the value of L, the latent heat, by an amount equal numerically to the difference between the dry- and wet-bulb temperatures. It would not be profitable to attempt to introduce such an amendment.

* B. Rolf. *Tables psychrométriques portatives. Calculées selon la formule Svensson-Ekholm* (Stockholm, 1919).

535.214

ON RADIOMETER ACTION AND THE PRESSURE OF RADIATION

By MARY BELL, formerly Lecturer in Physics,

AND

S. E. GREEN, Lecturer in Physics,

Imperial College of Science and Technology

Received July 1, 1932 *and in revised form December* 7, 1932. *Read January* 20, 1933.

ABSTRACT. The experiments here described have arisen from preliminary work conducted with the ultimate object of testing the late Prof. Callendar's formula* for the distribution of energy in the spectrum. The latter experiments, not yet completed, involved radiation-pressure determinations made with a vane suspended in a vacuum sufficiently high for radiometer action to be completely eliminated. While the manner in which radiometer action diminishes as the vacuum improves was being investigated, the series of radiometer effects recorded in the first part of this paper was observed. Radiation was directed upon a light vane suspended by means of a fine quartz fibre in a flask, and the deflections produced when the radiation was incident in turn upon the two sides of the vane were observed for a series of vacuum conditions. The vanes employed included platinized or silvered glass and mica, and also an aluminium vane. The range of air pressure used extended down to about 10^{-6} mm. of mercury.

In the case of the platinized glass vane, at a pressure of a few mm. of mercury the deflections on both the glass and the platinized sides were of the order of those registered at the highest vacua obtained. At pressures of about 10^{-2} mm. gas effects developed which were very large in the case of the glass and mica vanes but much less marked with the aluminium vane. With all the vanes, gas action had practically vanished at a pressure of 10^{-6} mm. of mercury. Under these conditions, measurements of radiation pressure, using glass and mica vanes, were found to give results in agreement with the energy-density of the radiation to within $\pm$ 7 per cent.

While the above preliminary experiments were in progress the errors of calculation occurring in Nichols and Hull's paper† on radiation pressure were noticed. Although their work, carried out at a gas pressure of 16 mm. of mercury, has been widely quoted as conclusively establishing, to within about 1 per cent, the numerical equivalence between the pressure and energy-density of radiation, it is found that their results, when correctly evaluated, show a divergence between these quantities of some 10 per cent. Hence Nichols and Hull's investigation cannot be regarded as furnishing a quantitative experimental verification of the equality relationship deducible from theory.

It was therefore thought worth while, using the working experience gained, to attempt more accurate measurements of radiation pressure by the direct method of using metal vanes and a vacuum so high that radiometer action was eliminated. In the second part of the paper the observations and results for all the radiation-pressure measurements are given. With each vane a considerable range of radiation-intensities was employed, the maximum deflections of the suspensions under the influence of the radiation-pressure being up to 10 or 15 times those obtained by Nichols and Hull. The corresponding energy-densities of the radiation were measured by the use of a Callendar cup radio-balance. The

* *Phil. Mag.* **27**, 870 (1914).

† *Proc. Amer. Acad.* **38**, 559 (1903).

results for an alloy vane, probably the most reliable of those derived, show that for a number of radiation-intensities the difference between pressure and energy-density varies from + 4 to − 3 per cent. The present experiments are considered in the light of those conducted by other observers. Errors of calculation arising in Nichols and Hull's results are examined in the appendix.

§ 1. PRELIMINARY

INTRODUCTION. It is well known that when radiation falls upon a vane suspended in an exhausted vessel the true radiation pressure is obscured, even when the vacuum is of the order of 10^{-4} mm. of mercury, by gas action produced by the heating of the residual air in the neighbourhood of the vane.

This so-called radiometer action has caused difficulty in verifying experimentally the theoretical relationship, originally deduced by Maxwell, between the pressure and the energy-density of the beam, viz. that these two quantities are numerically equal. Of the various attempts which have been made to circumvent gas action, the best known are those of Lebedew* and of Nichols and Hull†. The former made an effort to eliminate it by a suitable construction of the vanes, and the latter by the use of a ballistic method to which reference is made later in this paper. In only two investigations, those of Poynting and Barlow‡ and of Miss Golsen§, has the direct method of obtaining so high a vacuum that gas action becomes negligible been attempted.

Poynting obtained qualitative agreement between pressure and energy-density in the case of absorbing and reflecting surfaces; but, at the highest vacuum used by him, gas action was not entirely eliminated, nor did he investigate the changes in radiometer action at various stages of the vacuum. In 1924, when the work described in the present paper was in progress, Miss Golsen published her results for metal vanes, and showed that with these vanes a vacuum could be obtained at which gas action disappeared. The agreement she deduced between pressure and energy-density of radiation varied from + 6 per cent to − 4 per cent. Nichols and Hull state that their results agree to 1 per cent, but they make errors which are discussed at the end of this paper. Their results for pressure and energy-density, when properly calculated, differ by about 10 per cent.

The work described below was originally undertaken as a preliminary to a more delicate experiment, devised by the late Prof. H. L. Callendar, to test his formula for the distribution of energy in the spectrum of a black body‖. This experiment, which involves the measurement of the radiation pressure produced by different portions of the spectrum, and in which it is most important that radiometer action should be entirely eliminated, has not yet been completed. It has, however, been felt that the preliminary observations may be of interest in view of the fact that the investigation of radiometer action is more complete than that conducted by Miss Golsen; and also because, by a direct method, good agreement between pressure and

* *Ann. d. Physik*, (4), **6**, 433 (1901).
† *Proc. Amer. Acad.* **38**, 559 (1903).
‡ Poynting's *Collected Works*, p. 381; also *Proc. R. S.* A, **83**, 534 (1910).
§ *Ann. d. Physik.* **73**, 624 (1924).
‖ *Phil. Mag.* **27**, 870 (1914).

energy-density values has been obtained over a wide range of radiation-intensities, both with a platinized glass and with a metal suspension.

Oscillation of a vane suspended in a vacuum. In a large number of experiments on radiation pressure, including those described in this paper, the pressure is determined by allowing the radiation to deflect a light delicately suspended vane. Since in the present investigations the damping was so small that the vane was perpetually in oscillation, it is apposite to consider the mode of vibration of such a suspension. The equation of free motion of a vane suspended in a perfect vacuum can, on the assumption that any damping due to the suspending fibre itself is negligible, be written

t

$$\kappa \frac{\partial^2 \theta}{\partial t^2} + c\theta = 0,$$

κ, c, θ where κ represents the moment of inertia of the suspension and c the torsional constant of the fibre, while θ denotes the angular position at any instant and is given by the solution

A, α

$$\theta = A \cos\left(\sqrt{\frac{c}{\kappa}} . t - \alpha\right) \qquad \text{.....(1).}$$

P, r, L If radiation falling upon the vane exerts a total force P at an arm r, and the product Pr be written L, then the motion satisfies the equation

$$\kappa \frac{\partial^2 \theta}{\partial t^2} + c\theta = L,$$

of which the solution is

$$\theta = \frac{L}{c} + B \cos\left(\sqrt{\frac{c}{\kappa}} . t - \beta\right) \qquad \text{.....(2),}$$

B, β the constants B and β being determined by the values of the displacement θ and velocity $\partial\theta / \partial t$ obtaining at the moment of application of the radiation. Denoting

ϕ, v_0 these values by ϕ and v_0 respectively, and taking the instant of application of the light as the new zero of time, we find that

$$B = \{(\phi - L\,c)^2 + v_0^2\,\kappa\,c\}^{\frac{1}{2}}$$

and

$$\beta = \tan^{-1} \frac{v_0 \sqrt{(\kappa/c)}}{\phi - L/c}.$$

Relation (2) shows undamped harmonic motion about the mean displaced position $L\,c$, the position at which the couple due to the fibre balances the moment due to radiation pressure. The amplitude B of vibration about this mean position is reduced to zero only when the relations $\phi = L\,c$ and $v_0 = 0$ are together satisfied, i.e. when the amplitude of the initial free vibration of the suspension is $L\,c$, and the radiation is applied when the vane is at the extremity of its swing. The vane then remains steadily displaced in this position.

Thus, with a perfect vacuum, the vibration is perfectly regular about the mean displacement $L\,c$, the turning-points are invariable, and the periodic time is constant. When the light is removed, the vane immediately oscillates about its zero position, with constant turning-points and time of swing, the amplitude of the motion

depending on the position and velocity of the vane that obtained when the force was removed.

When the vacuum is imperfect, damping will arise owing to the presence of the residual air. If 2ϵ is the damping-factor, the equation of motion of the vane swinging freely is ϵ

$$\kappa \frac{\partial^2\theta}{\partial t^2} + 2\epsilon \frac{\partial\theta}{\partial t} + c\theta = 0,$$

with the solution

$$\theta = Ce^{-(\epsilon/\kappa)t} \cos(2\pi t/T - \gamma) \quad \text{.....(3);} \qquad \gamma$$

while when the radiation pressure is applied

$$\kappa \frac{\partial^2\theta}{\partial t^2} + 2\epsilon \frac{\partial\theta}{\partial t} + c\theta = L,$$

giving the solution

$$\theta = L/c + De^{-(\epsilon/\kappa)t} \cos(2\pi t/T - \delta) \quad \text{.....(4).}$$

Thus in both cases the time of swing T is $2\pi\sqrt{(c\kappa - \epsilon^2/\kappa^2)}$, while the amplitude of the oscillations decays exponentially, the logarithmic decrement of the motion being $\epsilon\pi/\sqrt{(c\kappa - \epsilon^2)}$. In practice, however, conditions are more complicated, since damping is not the only effect produced by residual air in the vacuum chamber. Experiments described later showed that, when on application of the light one surface of the vane was raised to a higher temperature than the other, pressures many times greater than the true radiation-pressure were generated. In this case the vane would suffer a mean displacement θ given by $(P \pm R)\, r = c\theta$, where R represents the pressure R produced by the heating of the residual air. Such pressures took time to develop and made the time of swing irregular. They also masked the regular decrease of amplitude of motion which would be expected from equation (4). A similar phenomenon was observed on removal of the light from the vane. In this case residual gas effects, which died away slowly, prevented the vane from vibrating about its true zero. These extraneous pressures, usually referred to as radiometer effects, were observed practically to disappear at a vacuum of about 10^{-6} mm. of mercury.

The objects of the experiments here described were firstly to study the variation of radiometer action as the vacuum improved; and secondly to investigate radiation pressure by obtaining so high a vacuum that the only forces acting on the suspensions employed were the true radiation pressure and the restoring elastic forces of the fibre.

As has already been indicated, in such a vacuum the vane will vibrate quite regularly about a mean angular displacement, and the turning-points and period of swing will be quite constant. When the light is removed the suspension will at once oscillate about its zero position, with constant turning-points and time of swing.

On attainment of these conditions it may safely be assumed that the true radiation pressure is being investigated. If, on the other hand, the turning points and time of swing of the vane vary and if, on removal of the light, the vane shows a lag in taking up its oscillation about the true zero, radiometer forces as well as the true radiation pressure are present.

The vanes used in the work to be described were, under experimental conditions, in continual vibration. For brevity, therefore, the term "deflection" is employed to denote the mean position about which the suspension is in oscillation.

§2. AN EXPERIMENTAL INVESTIGATION OF RADIOMETER ACTION

The apparatus employed is indicated in plan in figure 1. The optical system *p* was used to project a beam of light from the lamp *l* on to a vane *v* suspended within the flask *f*. This flask, 17 cm. in diameter, was connected through suitable apparatus, shown in figure 2, to the exhausting-system comprising a Gaede mercury-pump outfit.

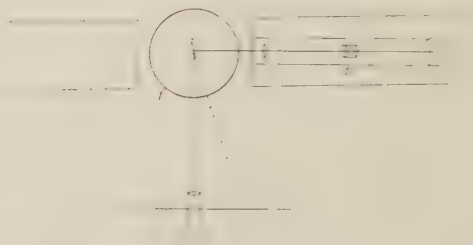


Figure 1. Plan of apparatus.

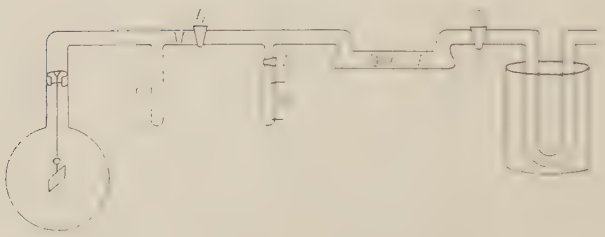


Figure 2. Flask and auxiliary apparatus: *C*, carbon tube; *G*, Geissler tube; *P*, pentoxide tube; T_1, T_2, T_3, taps; *U*, U-tube surrounded by liquid air.

Figure 3. The optical projection system.

When a very low pressure was required, both the flask and the carbon-filled tube were heated in a furnace to 360° C. during pumping. Figure 3 shows in elevation the arrangement of the source of light and of the lenses composing the projection system. The source was a 500-candle-power pointolite lamp. An image of the aperture of the photographic lens *B* was focused on the vane by the achromatic lens *A*. Of the wooden carrier, *EFG*, the portion *FG*, to which the lamp and lens

were attached, turned about a pivot P so that an image could be thrown on either end of the vane. As is indicated in figure 1, an appropriate baseboard was fitted on each side of the flask. By transfer of the lamp and projection system from one board to the other the radiation could be made to fall upon either surface of the vane as required.

The vanes were all rectangular in shape as shown in figure 4. In the case of glass vanes the centre piece AB was also of glass, fused to the vane. When a metal vane was used the vane and stem were cut as one piece from the sheet. The vanes were suspended by means of a fine quartz fibre attached at its upper end to a collar which fitted into the neck of the flask. The collar was of gas-free molybdenum supplied by the Vactite Company. The fibre was attached at each end by means of collodion. When this collodion had been gently heated by means of a very small gas flame a deposit of carbon remained which held the fibre well and did not affect the vacuum. Deflections of the vane were read by the lamp-and-scale method, the concave mirror c, figure 4, being used. By means of fine wire this mirror was bound to the stem of the suspension so as to lie at right angles to the vane. Throughout the experiments the scale-distance was of the order of 1 metre.

Details of the suspensions employed are given in table 1.

Table 1

Vane (*a*) Glass, platinized on one side
Vane (*b*) Glass, blackened on one side
Vane (*c*) Aluminium

Vane	Length (cm.)	Mass (gm.)	Moment of inertia (gm.-cm²)	Time of swing (sec.)
(*a*)	3·49	0·340	0·196	52·6
(*b*)	2·25	0·108	0·047	50·5
(*c*)	3·60	0·207	0·047	57·5

The thickness of the glass vanes was of the order of 0·25 mm. The aluminium vane was cut from a piece of foil about 0·05 mm. thick*.

Experiments with the glass vane platinized on one side. (1) To obtain some idea of the degree of vacuum reached in these and subsequent experiments, a McLeod gauge was attached, through a U-tube surrounded by liquid air, to the exhausting apparatus; and at various stages of the pumping the logarithmic decrements of the suspension and the corresponding pressures were noted.

Figure 5 shows the variation of logarithmic decrement with pressure. The values of the logarithmic decrement are given to the base 10. A decrement of 2×10^{-3} corresponded to the lowest pressure read on the gauge, i.e. 4×10^{-4} mm. of mercury.

(2) The McLeod gauge was disconnected from the apparatus and the pumping restarted. At various intervals tap T_1, figure 2, was turned, and the deflections

* It is to be emphasized that the effects described in this paper apply only to the case of flat rectangular vanes of the sizes indicated, suspended by means of a quartz fibre, centrally within a glass sphere about 17 cm. in diameter.

produced when a beam of constant intensity was incident in turn upon the glass and the platinized surfaces of the vane were observed. A constant arm was attained by making the edge of the vane a tangent to the circular image. In every case the mean of the deflections for the two ends of the vane was taken. The pressures corresponding to the various deflections were derived from the curve in figure 5. The lowest pressure obtained corresponded to a logarithmic decrement of 6×10^{-6} and was thus probably of the order of 10^{-6} mm. of mercury. In this experiment the flask was heated only to about 200° C., while in later work on radiation pressure the flask-temperature was raised to 360° and this, with the further use of a diffusion pump, makes it highly probable that still lower pressures were reached.

Experiments made with the radiation incident upon the platinized surface of the vane showed that all the deflections produced were in the direction of

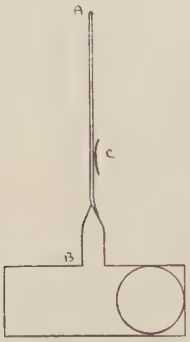


Figure 4.

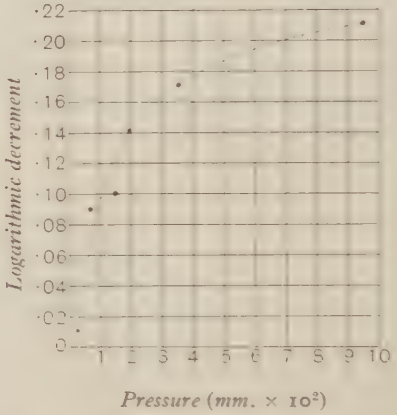


Figure 5. Variation of logarithmic decrement of the suspension with gas pressure.

pressure. At a vacuum of 10^{-2} mm. of mercury the deflection was very large, the spot of light which registered the movement of the suspension passing right off the scale. At 10^{-3} mm. of mercury the deflection had noticeably diminished. At a pressure of 10^{-6} mm. the deflection was reduced to about 30 mm. The deflection remained constant at this value during prolonged application of liquid air to the charcoal tube. It was found that the radiation pressure, calculated from this deflection, was in approximate agreement with the energy-density value. Since the logarithmic decrement of the motion of the vane was only 6×10^{-6}, damping was obviously negligible. The time of swing of the suspension under the radiation pressure was the same as during free oscillation, and the vibration was perfectly regular about the mean position.

With the radiation incident upon the glass surface of the vane, the deflections, which were originally in the pressure direction, became increasingly antipressure as the vacuum improved. This antipressure effect was very large when the gas pressure was roughly 10^{-2} mm. At about 2×10^{-3} mm., on application of the radiation, a pause

occurred before the suction effect began, while at 7×10^{-4} mm. there was a jump in the pressure direction and subsequently a suction effect. With a gas pressure of approximately 10^{-5} mm. of mercury the deflection was found to have changed sign; and it finally acquired a constant positive value (i.e. in the pressure direction) of about 30 mm. at a vacuum in the vicinity of 10^{-6} mm. of mercury. Figure 6 indicates in a general manner, for various gas pressures, the motion of the mean position of the suspension when the radiation was made incident upon the two sides of the vane in turn.

Similar experiments to those described above were conducted with mica vanes. The behaviour already indicated was found to be quite general for all cases in which glass or mica vanes, platinized or silvered on one side, were used.

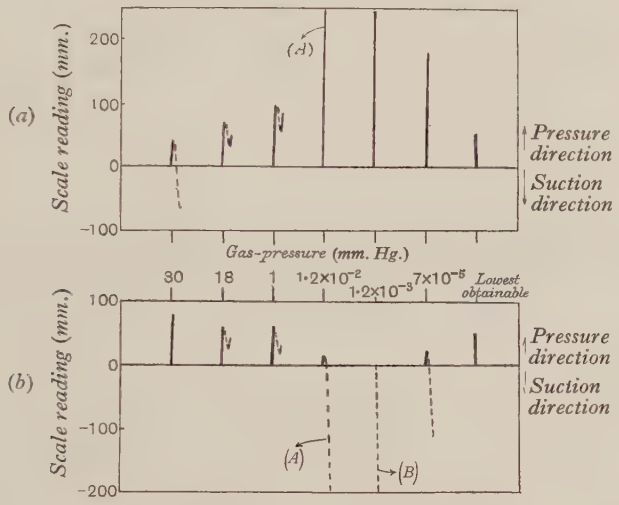


Fig. 6. Variation of deflection of platinized glass vane with gas pressure: (*a*) Radiation on platinized surface; (*b*) Radiation on glass surface*.

Notes. (*A*) Off the scale. (*B*) Off the scale, but with slight arrest before first movement.

Further experiments showing growth and decay of radiometer action at a vacuum of 10^{-4} *mm. Hg, and the disappearance of this action in the neighbourhood of* 10^{-6} *mm.* (1) After pumping to a vacuum of 10^{-4} mm. of mercury, tap T_1, figure 2, was turned, and the beam was allowed to fall on the glass side and kept there continuously. There was a jump of 25 mm. in the pressure direction, then a suction effect showing the growth of radiometer action (see figure 7). This radiometer effect reached a maximum fairly quickly and then decreased very slowly with time—see portion *MP* of the curve. This gradual decay of radiometer action with time when

* In this figure deflections in the suction direction are shown by dotted lines. At the pressures of 30, 18 and 1 mm. of mercury, the actual movements of the vane, before coming to rest, are given. The high degree of damping at these pressures is thus illustrated. At the lower pressures only the final mean position of the vane is indicated. The damping decreased rapidly after about 10^{-1} mm. and was negligible at 10^{-6} mm.

light was continually on the vane and the vacuum was constant occurred whether the radiation was directed on the glass or on the platinized side; that is to say, the deflection apparently tended ultimately to become the true pressure deflection. This rate of decay was so extremely slow at pressures of 10^{-3} to 10^{-5} mm. that the

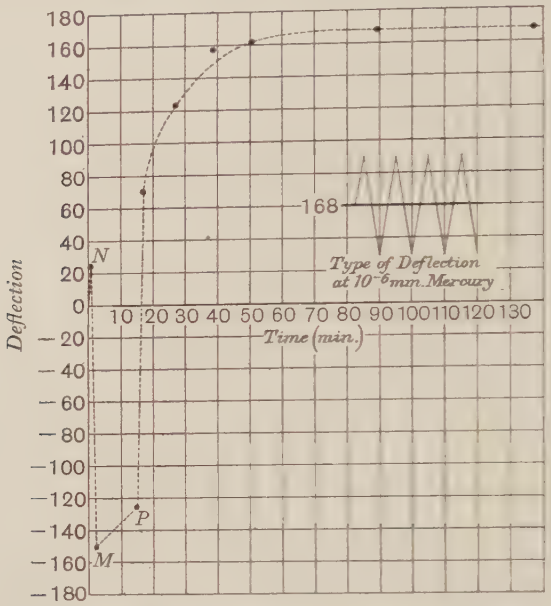


Figure 7. Illustrating radiometer action with light on glass surface, at gas pressures 10^{-4} to 10^{-6} mm. Hg.

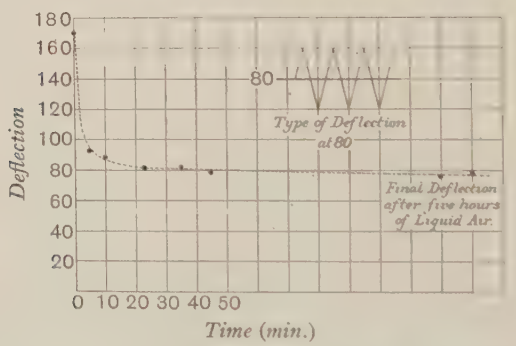


Figure 8. Illustrating radiometer action with light on platinized side at gas pressures from 10^{-4} to 10^{-6} mm.

possible complete disappearance of radiometer action, which might be inferred, was never actually observed. As will be seen later, the effect at higher vacua was much accelerated.

Without the light being removed from the vane the vacuum was now brought down to about 10^{-6} mm. by applying liquid air to the carbon-filled tube. Following

the change of vacuum the deflection quickly became a pressure effect, and the vane eventually oscillated regularly about a fixed mean position during prolonged application of liquid air. This is shown in the inset figure, where the new time scale is such that the distance between successive maxima is the period of the suspension (about 53 sec.). Although denoted by straight lines for convenience, the oscillations were in fact simple harmonic, and the vibration is of the exact type given by equation 2, page 322.

(2) A similar experiment was carried out with radiation incident upon the platinized side of the vane (see figure 8). In this case an initial deflection of 170 mm., produced by applying the radiation at a vacuum of 10^{-4} mm., was reduced in 25 minutes to a steady deflection of 80 mm. on the application of liquid air to the carbon-filled tube. The turning-points at this highest vacuum were regular during prolonged application of liquid air, as shown in inset figure 8, where all considerations mentioned above regarding inset figure 7 again apply.

Radiation was next directed upon the glass and platinized sides respectively at a vacuum of 10^{-6} mm. Table 2 shows the observed deflections in millimetres of scale.

Table 2

Radiation on platinized side			Radiation on glass side		
Turning-points		Deflection from zero	Turning-points		Deflection from zero
L.	R.		L.	R.	
54	203	75	228	87	70·5
57	195	69	236	83·5	76
56	195	69·5	237	82·5	77
56	193·5	68·5	238	82·5	77·5
56·5	193·5	68·5			

A study of the deflections shows that even at this high vacuum radiometer action existed initially to some extent. The quick growth (see page 327, line 18) is not here apparent, as it is masked by the relatively long period of the suspension, but the presence and the decay of radiometer action can be detected. On the platinized side the deflections diminish until they reach a steady value, i.e. on the platinized side the gas action exists initially as a pressure and gradually disappears. On the glass side the deflections rise until they reach a steady value, i.e. the radiometer action is an anti-pressure effect and gradually disappears. Ultimately the deflections have constant turning-points, the time of swing is constant, and the logarithmic decrement is negligible, i.e. in every way the oscillations conform to equation (2).

It may be objected that even in these circumstances the entire absence of radiometer action is not proved. Some constant gas action might conceivably be present in conjunction with the true radiation pressure. From radiation-pressure experiments with an aluminium vane, strong additional evidence was, however, forthcoming that when the vibrations of this aluminium vane conformed exactly to equation (2), radiometer action was entirely absent. It seems therefore reasonable to

assume that if the same equation is satisfied in the case of the glass suspension, gas action has been completely eliminated.

One further effect was observed. The final deflections on the platinized and glass sides were not equal, but the former were always less than the latter as illustrated in table 3. Incidentally this table also illustrates for various radiation-intensities the decay of radiometer action just described.

The difference between the final deflections on the two sides is thus of the order of 16 per cent. The data given in the table were obtained during experiments upon the pressure of radiation described later*. The mean reflecting power (see table 8) on the platinized side of the vane was 67 per cent and upon the glass side 65 per cent. It is therefore impossible to explain the recorded divergence between the final deflections on the basis of a divergence in the reflecting power on the two sides. Again, radiometer action, as has been shown, always tends to increase the deflection on the platinized side and to decrease that on the glass side. The divergence between the final deflections cannot, then, be put down to residual radiometer action unless, as seems very improbable, this changes sign at very low pressure. Incidentally there seems to be here further support for the view that radiometer action has vanished at these very low pressures.

Table 3

Beam	Platinized side		Glass side	
	Deflections		Deflections	
	First	Final	First	Final
1	48·0	37·6	34·1	44·2
2	78·1	65·4	72·9	77·2
3	96·1	88·2	93·2	102·7
4	118·0	112·6	123·2	126·6
5	132·0	116·7	132·2	140·2
6	155·5	150·6	160·2	170·6

The suggestion is put forward that the effect may be due to the fact that the radiation absorbed by the film (some 35 per cent of the incident radiation) is re-radiated on one side through glass and on the other through a vacuum.

Assuming that this absorbed energy is reradiated on the glass and vacuum sides in the ratio of $\mu^2 : 1$, taking a tentative value of μ as 1·5, and assuming also that this reradiated energy is not absorbed by the glass, we find that the correction for the reaction on the vane due to reradiation reduces the disparity between the deflections from some 16 per cent to about 5 per cent, while leaving the final mean value unaltered.

Figure 9 shows in a very general manner, and not to scale, the deflections of the suspension when the radiation was kept permanently on the vane as pumping continued. It will be noticed that at a pressure of a few millimetres of mercury

* A second platinized glass vane was used for these later experiments. Data relating to this suspension are given, in connection with the radiation pressure work, on p. 338.

the deflections for both the glass and the platinized surfaces are approximately equal to the true radiation deflection. These curves were taken for three different radiation-intensities and are of interest in that they appear to corroborate Nichols and Hull's observation that, with a gas pressure in the vicinity of a few millimetres of mercury, the resultant of combined gas actions has so small a value that the deflection of the vane is of the same order as the true radiation deflection. Owing to the rapid change of vacuum obtaining in these experiments, it must be emphasized that no deductions other than that just mentioned can be made with regard to gas action.

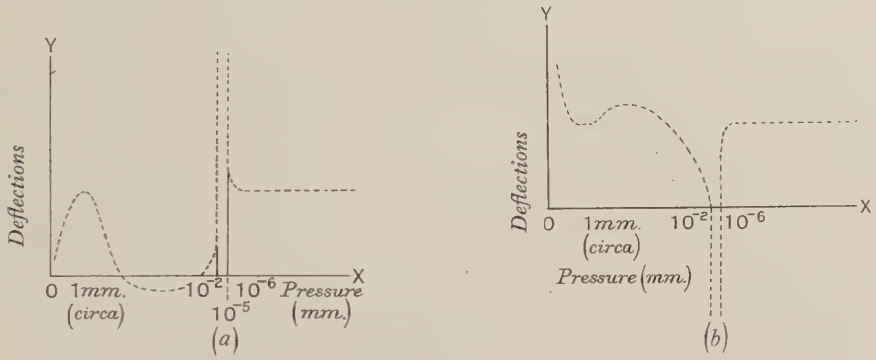


Figure 9. (*a*) Platinized surface. (*b*) Glass surface. Variation of deflections with the light continuously on the vane as pumping continued.

Summary. The experiments already described are typical of a large number conducted with silvered and platinized glass vanes. In every case the following general results were established:

(*a*) At a pressure of a few millimetres of mercury the deflections on both the glass and the platinized sides were of the order of those registered at the highest vacuum obtained. (*b*) At pressures of about 10^{-2} mm. large gas effects developed and were apparent as a suction on the glass side and as a pressure on the platinized side. If D_P denote the deflection due to radiation pressure, D_R the deflection due to gas action and D the resultant deflection, then $D_R + D_P = D$ where D_R is negative for the glass surface and positive for the platinized surface. (*c*) D_R took time to develop, reached a maximum and then tended to die away very slowly. (*d*) The higher the vacuum the less was the gas action, and D_R eventually became less than D_P. When this occurred D was positive on both surfaces. (*e*) At a vacuum of the order of 10^{-6} mm., when D_R was very small no initial growth of gas action was apparent, but the much slower decay was easily observed. The mean value of the final steady deflections was now taken to give the true radiation pressure. (*f*) The final steady deflection on the glass side was always greater than that on the platinized side.

D_P, D_R
D

Experiments with the glass vane blackened on one side. In the case of this vane the flask was heated to 360° C. during pumping, and the carbon tube was immersed in liquid air for a long time prior to the application of the radiation. Even with

these precautions there was a large amount of residual gas action, shown by the fact that the initial deflection on the glass side was very much smaller than that on the blackened side, and the deflections only became equal after an appreciable time interval. Typical data are as follows:

Table 4. (Light continuously on the vane)

	Initial deflections	Final deflections
Glass side	27	42
Blackened side	64	42

The state of steady vibration about the final mean position 42 was reached after 10 oscillations in each case.

It was further noticed in this experiment that if the light was removed after a steady deflection had been attained the vane did not at once take up its zero position. There always appeared to be a slight residual suction effect when the light had been incident upon the blackened side and a slight residual pressure effect when the radiation had been falling upon the glass surface. These deflections gradually died away as is shown in the following typical table.

Table 5

	Steady-pressure deflection	Residual deflections after removals of light
Glass surface	35	(Pressure) 5·5, 2·5, 1·5, 1·0
Blackened surface	42	(Suction) 9·0, 5·5, 3·5

The above table shows that with residual radiometer action the vane does not at once take up the original zero when the radiation is removed.

The time interval separating the consecutive deflections given in the third column was about 51 sec., the period of the suspension. It may be said that, with this vane, gas action was in every way similar to, but very much larger than, that with the platinized vane.

Experiments with the aluminium vane. When the vacuum was of the order of 10^{-6} mm. the vane assumed its final condition of steady vibration directly the beam of light was allowed to fall upon it (compare the residual gas effects with glass vanes already mentioned) and fell to vibration about its original zero immediately the light was removed.

In Table 6 are given deflectional observations in millimetres of scale with beams of different intensities with a vacuum of about 10^{-6} mm. pressure.

These observations illustrate the steadiness of the zero and also of the turning-points.

When the vacuum was inferior the deflection of the vane did not at once assume its final value, and the zero wandered.

Table 6

Beam	Zeros		Turning-points with beam on vane		Deflection
	Before deflection	After deflection			
	Left	Left	Left	Right	To the right
I	43·5	44	79·5 79·5	o 58·5 58·5	33·2
II	43·7	43·7	85 86·5	o 80 80	40·8
III	43·7	43·7	70 69·6	o 81·5 81	49·4
IV	43·7	43·4	77·5 77·4	o 110 110	59·9
V	43·4	43·4	17·5 17·5 17·5	o 42·8 43·4 43·2	73·7

General remarks on the observed effects. Since the platinum film reflected 65 per cent only of the incident light while producing practically no scattering, it absorbed 35 per cent of the radiation falling upon it. The glass plate was found experimentally to absorb about 4 per cent of the incident light. The mass of the platinum film was roughly one-fifth of that of the glass plate. The specific heat of the film (assuming this constant to be the same for the sputtered material as for the substance in bulk) was 0·03, that of the glass 0·2. Hence the thermal capacity of the platinum film was much less than that of the glass plate. It follows that the film always rose to a higher temperature than the glass plate, whether the light fell upon the glass or upon the platinized surface of the vane. Calculation showed that the temperature-difference between the surfaces was of the order of 0·1° C. In accordance with the usual explanation* the velocity of recoil of the molecules meeting the hotter surface must be greater than the velocity of approach. This gives rise to a pressure on the heated surface, and strictly speaking this pressure only is referred to as "radiometer action." Thus, radiometer action will always be a pressure on the platinized surface of the vane, whether the radiation is incident upon the platinized surface or upon the glass surface. Consequently the radiometer effect will be in the same direction as the radiation pressure when the light is incident upon the platinum film, and in the opposite direction when it falls on the glass side; i.e. there will apparently be a pressure effect in the former case and a suction effect in the latter.

At the higher gas pressures it is probable that convection is an important factor. Reference to figure 6 shows that the total gas action reaches a maximum at pressures of 10^{-2} to 10^{-3} mm., and from the directions of the deflections it looks as though

* Poynting, *Heat*, pp. 150, 151 (1922).

radiometer action had now become the only gas effect. The very large value at these pressures may possibly be connected with the fact that the mean free path of the residual gas molecules is now of the same order as the dimensions of the vane. Obviously radiometer action must become negligible at a sufficiently high vacuum.

In his discussion of radiometer action, Poynting* states that over the surface of the vane the extra pressure produced by the heated molecules would be quickly compensated by an adjustment of density of the gas, but that the compensation is not complete at the edge of the vane, so that persistent radiometer action is really an edge effect. Experiment showed that at pressures of 10^{-2} to 10^{-4} mm. radiometer action died away extremely slowly, though it decayed quickly at lower pressures, possibly owing to the greater mobility of the gas. All the experiments recorded, however, were performed with vanes not materially differing in size or shape and with air or oxygen as the residual gas. Thus the results given apply only to these particular conditions and throw no light on the existence of a possible edge effect, or on the influence of the nature of the residual gas. If it were possible to assume that the density compensation at the surface of the vane mentioned by Poynting proceeded very slowly at 10^{-2} to 10^{-4} mm. and more quickly at 10^{-6} mm., the gradual fall in deflections on the platinized side and the gradual rise on the glass side until the steady state was attained, and also the lag in returning to the true zero, when the radiation was removed, could be explained without reference to an edge effect. In any case it is probable that the lag effects observed are connected with the time taken for the entire system to acquire the thermal steady state.

There was found to be an absence of regular gas effects with the aluminium vane. Such effects as were present were very small, had a maximum value at about 10^{-3} mm., and had entirely ceased at a pressure of 10^{-6} mm. This was to be anticipated in view of the high thermal conductivity and thinness of the vane. The two surfaces of the latter would be at practically the same temperature, so that no perceptible gas action could arise. Miss Golsen also found maximum radiometer effects at a pressure of about 10^{-2} mm. and states that at a pressure of 10^{-6} mm. radiometer action ceased with a metal vane.

All the vanes were heated to 300° C. in a vacuum before being mounted in the flask. For this reason, and also because the radiometer effects were extremely regular, it is unlikely that they can have been in any way due to escape of gas from the surfaces of the vanes during the experiments. All the results mentioned were perfectly definite and reproducible.

§ 3. DETERMINATION OF THE PRESSURE OF RADIATION AND THE COMPARISON OF THIS PRESSURE WITH THE ENERGY-DENSITY

Apparatus and general procedure. For these measurements the apparatus employed consisted of the following: (1) a light vane of glass or metal suspended in an exhausted flask; (2) an optical system comprising a gas-filled Osram lamp and

* Poynting, *Heat*, 150, 151 (1922).

suitable lenses so arranged that the radiation from the lamp could be directed upon either of the surfaces of the vane; and (3) a Callendar cup radio-balance* to measure the intensity of the beam and the reflection coefficient of the vane.

The arrangement of the apparatus and the procedure followed in the evacuation of the system were the same as in the investigation of radiometer action (see figures 1, 2 and 3) with the two exceptions that the Gaede mercury pump was eventually replaced by a mercury diffusion pump backed by the Gaede pump, and that, with the object of improving the vacuum, the carbon tube and flask were heated to higher temperatures during evacuation. A good vacuum was obtained when flask and tube were heated to 360° C. for eight hours or more during pumping; the temperature of flask and carbon tube were then slowly reduced to normal while pumping continued. Taps T_1, T_2 and T_3, figure 2, were then closed in order, and the furnace was removed. An asbestos case was placed round the flask, and when the vane had lost the extraneous motions produced by these operations the carbon tube was surrounded by liquid air and observations were taken. In the earlier experiments the flask was sealed off from the exhausting apparatus by means of an electric furnace surrounding the arm M, figure 2. This operation appeared, however, to be unsatisfactory, for two reasons. In the first place gas was emitted from the red-hot glass and thus the vacuum was impaired, and secondly great risk was involved of spoiling the experiment by some accident. Specially prepared tap grease was used for T_1, figure 2. From the excellence of the vacua obtained after the carbon tube had been surrounded by liquid air, it was obvious that the vapour pressure of the grease had a negligible effect. A further great advantage of this arrangement was that a series of evacuations could be performed without disturbing the flask and the suspended system within it. The vanes used were heated to 360° C. in a vacuum before being mounted in the flask.

In adjusting the position of the suspension within the flask two conditions had to be satisfied. The vane had to lie at right angles to the direction of the beam, and at an elevation within the flask such that no radiation was reflected back from the flask surface to the vane. The image produced by the flask of the focused spot on the vane was readily seen within the flask, and the vane was so adjusted that this image fell just below it. After the necessary adjustments had been carried out by means of the collar in the neck of the flask the latter was sealed to the exhausting apparatus.

For some experiments the quartz fibres employed were made by the bow-and-arrow method; but the more delicate ones were drawn out by means of an oxy-hydrogen blowpipe by a method shown to the authors by Dr W. F. G. Swann. These very delicate fibres showed colour in sunlight. The finest ones, of a blue colour, were not strong enough to support the suspension. Those exhibiting a red colour were used. The fibres were attached to suspension and collar by a fine carbon layer by the method already described.

The 1000-watt filament lamp used as the source of light was connected to a steady 110-volt supply, the potential-drop across the lamp being maintained con-

* *Proc. Phys. Soc.* **23**, 15 (1910).

stant by means of a resistance in circuit. The lamp was found to remain very steady over long periods.

Calculation of the radiation pressure. The following system of symbols is employed:

m — m is the reflecting power of the vane;

P — P the required pressure of the radiation;

A — A the area of the spot of light on the vane;

r — r the arm at which the radiation pressure acts;

c — c the fibre constant, i.e. the couple necessary to turn the suspension through unit angle;

θ — θ the mean angular deflection of the vane;

T — T the periodic time of the suspended system; and

I — I the moment of inertia of the suspension (vane and concave mirror).

The pressure of the beam was deduced from the relations

$$(1+m)\,PAr = c\theta \quad \text{and} \quad T = 2\pi\sqrt{\frac{I}{c}} \qquad \text{.....(5) and (6),}$$

which give

$$P = \frac{c\theta}{(1-m)\,rA} = \frac{4\pi^2 I d}{T^2 rA\,(1-m)\,2D} \qquad \text{.....(7) and (8),}$$

d, D — where d and D denote respectively the scale deflection and the distance between the scale and the mirror on the stem of the suspension. Correction of the value of $d/2D$ to give the true angular displacement of the suspension was in most cases too small to justify application. For the same reason, correction for the fact that, in the displaced orientations of the vanes, the radiation was not quite normally incident could in general be disregarded.

The lens system was so arranged that the ends and two edges of the vane could be made tangents (see figure 4) to the focused image of the aperture of lens B, figure 3. Thus from the known dimensions of the vane the area of the image (A above) and the arm r at which the light-pressure acts could be calculated from the relations

$$A = \pi\,(\tfrac{1}{2}b)^2 \quad \text{and} \quad r = \tfrac{1}{2}\,(l-b) \qquad \text{.....(9) and (10),}$$

where l and b denote respectively the length and depth of the vane.

In some cases the area of the spot was deduced from the aperture of the photographic lens B, figure 3, and the appropriate constants of the optical system. The deflection d was determined for each end of the vane as well as for both surfaces.

Determination of the reflecting power of the vane. The measurement of the reflection coefficient of the vane was performed in a subsidiary experiment, the incident and reflected intensities being observed by means of the radio-balance*. The optical arrangement employed for the production of direct and reflected beams was similar to that adopted by Nichols and Hull, and is shown diagrammatically in figure 10. In this diagram M is a concave mirror, RB the radio-balance and

* The method of deriving the radiation-intensities is indicated on p. 338.

V the vane fixed in an upright at the centre of the turntable T. The full line indicates the path of the radiation when directly incident upon the concave mirror and radio-balance, and the dotted line the direction of the beam when reflection takes place at the surface of the vane.

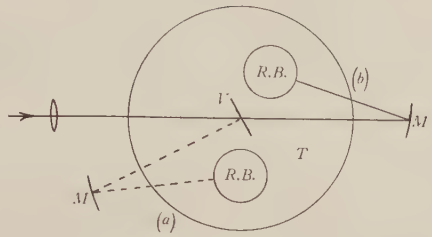


Figure 10. Indicating the method of measuring the reflecting powers of the vanes.

The procedure was first of all to cast a sharply focused spot of light on to the vane by means of the projection system used in the deflectional observations. An image of this spot of light was cast on to the radio-balance by placing the concave mirror and balance in position *a*, figure 10. In this position reflected radiation-intensities were measured. The incident intensities were determined after removal of the vane and rotation of the turntable into position *b*. The diameter of the spot of light used was varied, as required, by the use of a series of different apertures for the photographic lens B, figure 3. In preliminary experiments on reflecting power a number of types of thermopile were tried. Owing to the large uncertainties in the results arising from parasitic effects in the thermocouple circuits, it was necessary eventually to abandon the thermopile as the measuring instrument in favour of the radio-balance, which gave unambiguous and reproducible readings.

Since the vane was used in a vacuum, while the reflecting power was measured in the air of the laboratory, the assumption is involved that the reflection coefficient would not be substantially different in the two cases.

Measurement of the energy-density of the radiation. In determining the energy-density of the beam, previous investigators have had recourse to the indirect method of measuring the heat absorbed when the radiation was incident upon a given mass of a substance of known specific heat. Over such methods, which involved long experiments and complicated apparatus, the use of a Callendar radio-balance carried the great advantage that measurements could thereby be rapidly made to give the intensity of the radiation directly in watts per square centimetre. The latter was deduced from the relation

$$aH = 2PC + (D - D')/s \qquad \text{.....(11)},$$

in which a is the area of the balance cups, H the intensity of the beam, P the Peltier coefficient of the thermocouple junctions at the temperature indicated by a thermometer mounted in the instrument, C the neutralizing current and s the sensitivity of the galvanometer. D and D' denote the small deflections of the galvanometer obtaining when the radiation falls into the balance cups in turn.

a, H, P

C, s

D, D'

The term containing these deflections in the above expression constitutes a small correction.

If aH be the value in watts of the energy received per second by the balance, the energy-density of the radiation is

$$\tfrac{1}{3}H \times 10^{-3} \text{ erg/cm}^3$$

The spot of light focused on the radio-balance, being of the same size as that on the vane during the deflectional observations, was considerably larger in diameter than the balance apertures. The mean was therefore taken of a number of readings for different portions of the spot. Similar considerations applied to the determinations of the reflection coefficient of the vane.

For the evaluation of the intensity of the radiation falling upon the vane the radio-balance results required correction to allow for the energy loss due to the flask. The determination of this correction is considered later.

Measurements with a platinized glass vane. A series of determinations were made with a glass vane platinized on one side. The vane was exactly similar to that used in the later parts of the investigation of radiometer action (see page 329).

Specimen data and calculations involved in one typical determination of the radiation pressure with this vane are as shown in tables 7 and 8, and below.

When measuring the reflecting power of the vane, a large number of readings were taken, as the results varied to a considerable extent and the reflection coefficient constituted a large correction in evaluating the radiation pressure. A rather small spot of light was used in these determinations. The data given below show that the reflecting power was fairly constant over the right-hand side of the platinized surface and both sides of the glass surface, but varied over the remaining part of the platinized surface. Some of the figures given in table 8 were mean values taken from several readings.

Table 7. Deflections (mm.) of vane under the influence of the radiation
(Scale distance $D = 110{\cdot}7$ cm.)

Beam incident upon	Left-hand-side turning-points		Deflection from zero	Right-hand-side turning-points		Deflection from zero
	Left	Right		Left	Right	
Platinized surface	181·5	o 18	81·7	54	o 203	74·5
	154	24·5	64·7	57	195	69·0
	150	26·5	61·7	56·5	195	69·5
	151·5	25·5	63·0	56·5	193·5	68·5
	151·5	27	62·2	56·5	193·5	68·5
	151·5	27	62·2			
Glass surface	228·5	o 87	70·7	18	o 168·5	75·2
	236·5	83·5	76·5	18·5	169·5	75·5
	237·5	82·5	77·5	17·5	170·5	76·5
	238·5	82·5	78·0	17·5	169	75·7
	237·5	81·5	78·0	16·5	170·5	77·0
	238	82	78·0			
			Mean deflection on platinized side			65·4
			Mean deflection on glass side			77·2
			Final mean			71·3

Table 8. Reflecting power (per cent)

Platinized surface		Glass surface	
Left-hand side	Right-hand side	Left-hand side	Right-hand side
68·0	72·6	64·2	63·8
62·2	71·4	64·2	67·7
57·5	70·6	63·2	66·6
58·0	70·0	62·1	65·6
63·7	69·6	65·4	67·3
67·5	69·0		
64·3	70·8		
62·2	71·4		
—	73·1		
Mean values:			
62·9	70·7	63·8	66·2
Final mean value of reflecting power 65·9			

Calculation of the light-pressure:

Area of the spot, $\pi\left(\frac{1{\cdot}3087}{2}\right)^2$ cm.^{2}

Period of the suspension, 67·2 sec.
Deflection, 71·3 mm.
Arm, 0·933 cm.
Moment of inertia of suspension, 0·1485 cm.^{2}
Scale distance, 1107 mm.
Reflecting power of vane, 65·9 per cent.

Hence from relation (8) on p. 336,

$$P = 20{\cdot}1 \times 10^{-6} \text{ dyne/cm.}^2$$

Measurement of the energy-density of the beam. After each deflectional observation the optical system was turned so as to direct the radiation upon the radio-balance, and a series of energy-determinations were carried out. On the completion of these sets of experiments the reflection coefficient of the vane and the energy-loss due to the glass of the flask were measured. The latter was found by allowing the radiation to fall upon the radio-balance (*a*) directly and (*b*) through the flask. A number of readings for the energy received by the balance were taken, and it was found that the mean value of the loss due to the two sides of the flask was 21·8 per cent, giving for one side 11·6 per cent*.

The mean value of the energy falling per second upon the radio-balance apertures of area *a* when the radiation was directly incident was 2213 μW. Reducing this by 11·6 per cent we have 1956 μW. as the value of the energy falling

* Tests conducted with the flask (*a*) exhausted to about 0·1 mm. and (*b*) containing air at atmospheric pressure gave no systematic divergence in results. Thus any effect due to the presence of air between the flask surfaces must, if existent at all, have been small. This justified the later adoption of the more direct method of determining the energy density subsequently described.

on the vane per second per 0·03467 cm.^{2} Thus the energy-density of the beam exerting a pressure of $20{\cdot}1 \times 10^{-6}$ dyne/cm.^{2}, i.e.

$$18{\cdot}8 \times 10^{-6} \text{ erg./cm.}^3$$

Since the flask was large and the light entered practically normally at the equator, no appreciable distortion effect due to the transmission through the flask could be observed. In later experiments any possible distortion was automatically corrected by taking all the energy-determinations after the pressure-deflections. The piece of glass through which the light had passed was cut from the flask and placed in front of the radio-balance at exactly the same distance as it had been from the vane. Thus the optical conditions obtaining during the energy-determinations reproduced exactly those of the deflectional observations. All subsequent determinations by this more accurate method with flasks similar to the one referred to above gave an energy-loss due to the flask of about 9 per cent. It therefore seems reasonable to suppose that the result, 11·6 per cent, was about 3 per cent too high. This may in part account for the fact that in this experiment (see figure 11) the energy-density was always less than the pressure.

Further determinations with the platinized glass vane. A series of similar measurements were made with this suspension. Various values of light-pressure and corresponding energy-density of beam were obtained by placing the lamp in different positions relative to the rest of the optical system. The mean values of the deflections on the platinized and glass surfaces are given in table 9.

Table 9. Deflections (mm.) of the vane under the influence of the radiation

Beam	Platinized surface	Glass surface	Mean
1	37·6	44·2	40·9
2	65·4	77·2	71·3
3	88·2	102·7	95·5
4	112·6	126·6	119·6
5	150·6	170·6	160·6

Corresponding energy determinations are as follows:

Table 10. Radiation-intensities (W. $\times 10^{-5}$ per 0·03467* cm.^{2}) with beam directly incident upon balance

Beam	1	2	3	4	5
Intensity	124	231	309	363	501
	121	221	305	374	515
	126	207	295	365	487
	133	221	299	368	496
	—	226	312	376	501
Mean	126	221	304	369	500
Less 11·6 %	111	195	269	326	442

* Aperture of radio-balance cups.

The value (11·6 per cent) of the correction for the loss due to the flask which is mentioned both in the above table and in the earlier specimen calculation was the mean figure deduced from experiments with a variety of radiation-intensities.

Experiments with a mica vane. A set of similar experiments were conducted with a mica vane platinized on one side.

The final results deduced with the glass and mica vanes are plotted in figure 11.

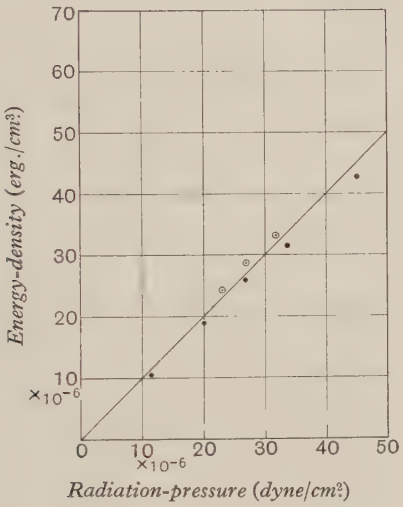


Figure 11. Results obtained using glass and mica vanes.
⊙ Mica vane • Glass vane

Later determinations with metal vanes. For various reasons the results given above for the platinized vanes were not considered to show the highest degree of accuracy obtainable by this direct method. They did however demonstrate that radiation pressure could be measured by it, and that therefore the experiment devised by Prof. Callendar could be proceeded with. This had been originally the fundamental issue. In the meantime, however, the errors occurring in Nichols and Hull's calculations had been noticed; it therefore seemed worth while to use the experience already gained to attempt more accurate measurements of pressure and energy values.

In this new experiment the material of the vane was aluminium. This metal was selected for its lightness and high thermal conductivity, and for the fact that it was already known that with a vane of this material radiometer action had entirely disappeared at a vacuum of 10^{-6} mm. of mercury. Various further points had to be considered. Since the pressure to be measured was extremely small, it was essential that a delicate fibre be used. Again, an important point was to find a workable value for the period of the suspension. If the period is too long, not only do the observations become tedious but the source of light may vary during the experiment. It was found that the end points could be accurately read and other conditions satisfied when the period was about one minute. The sensitivity

of the suspension could not profitably be raised by increasing the length of the vane, and with it the arm r, since to do so would unduly increase the moment of inertia and hence the time of swing. The experimental setting of the spot into the correct position might appear to be one of the most doubtful features of the work, but an examination of tables 11 and 13, which give readings by different observers, will show that this setting can be made with an accuracy not inferior to that obtained in other parts of the experiment.

Determinations with an aluminium vane. The series of observations conducted with an aluminium vane were carried out in a manner similar to that adopted in the cases of the non-metallic vanes previously described. Typical readings for the deflection of the suspension under the action of the radiation pressure are given in table 11.

Table 11. Deflection of the suspension (mm.)

Zero positions		*Turning-points with radiation on		Deflection
Before deflection	After deflection	left-hand side of vane	right-hand side of vane	
Radiation-intensity 1				
6·0	7·2	76·5 41·5 0	37·0 0 195·0	68·9
		75·5 40·0 0	36·5 0 197·5	
		75·4 39·8 0	37·5 0 197·5	
7·2	7·0	80·0 39·0 0	38·5 0 197·5	69·1
		78·5 38·0 0	38·0 0 198·0	
		78·0 38·0 0	38·0 0 198·0	
		78·8 38·5 0		
7·0	7·5	69·0 50·0 0	46·0 0 203·0	67·0
		68·0 49·0	46·0 202·0	
		67·8 48·5 0	46·0 202·0	
7·5	7·5	78·0 40·5 0	0 60·0 96·5	68·8
		78·0 41·0	60·0 97·0	
7·0	—	83·5 35·5 0	0 50·0 103·4	68·1
			0 50·0 103·4	
Radiation-intensity 2				
6·5	—	106·0 37·5 0	0 63·5 112·5	78·8
		105·5 37·5	64·0 113·0	
		208·0 0 67·0	0 46·0 132·0	80·0
		208·5 67·0	46·0 132·3	
		130·5 12·0 0	0 70·0 106·5	79·7
		130·5 11·6 0	70·0 107·0	
7·2	6·7	133·0 11·5 0	0 67·0 110·0	80·2
		132·0 11·5	67·0 110·0	

For the above and other intensities of radiation employed, the steadiness of the zeros of the suspension and the constancy of the amplitude of swing showed that gas action had been eliminated.

* It is to be remembered, that, with a fixed incident beam, while the mean deflection is constant, the amplitude of swing about this mean position varies with the velocity and position of the vane obtaining when the light was applied (see page 322). The amplitude, when once assumed, will then remain invariable.

With this vane two experiments were performed by evacuating the system twice, all conditions with the exception of the distance of the scale from the mirror being exactly reproduced.

A summary of the deflections and the energy-density values is given in table 12.

Table 12

Energy-density (erg/cm³)	Deflection of vane (mm.)
Evacuation 1. Scale distance 102 cm.	
$12{\cdot}2 \times 10^{-6}$	68·4
14·5	79·6
21·6	116·2
30·1	166·3
Evacuation 2. Scale distance 123 cm.	
$10{\cdot}8 \times 10^{-6}$	72·0
15·4	102·1
24·0	160·7
36·2	243·8

Unfortunately the surface of the aluminium changed during the determinations of the deflections of the vane, owing perhaps to the high temperature to which the metal was subjected or to the presence of a trace of mercury during the evacuation. Light on leaving the vane was found to be so scattered that the reflection coefficient could not be measured accurately. As nearly as could be estimated the value came to 42 per cent. With this value it was found that the calculated values for pressure and energy-density of radiation differed consistently by about 12 per cent. The pressures plotted against the energy values are given in figure 12. It will be seen that the results for the two evacuations lie closely on the same straight line, which on production passes through the origin. The fact that the ratio of calculated pressure to energy-density of radiation, while differing appreciably from unity, was thus constant both for varying intensities of radiation and for two independent evacuations, may be regarded as affording conclusive evidence of the absence of radiometer action. This is further supported by the regularity of the turning-points and zeros of the suspension illustrated in table 11*. It may be remarked here that with the platinized glass surface there was practically no scattering, so that as far as optical properties were concerned it was found to be greatly superior to all metal vanes—with the possible exception of the alloy vane used later.

Determinations with an alloy vane. Since the surface of the aluminium vane was found to be unsatisfactory, tests were made as to the suitability of other metals for the vane material. The metal finally adopted was an alloy consisting largely of

* Thus while it was not possible to deduce the absolute value of the pressure of radiation, the observations taken with the aluminium vane were of great interest for two reasons. They showed conclusively that radiation pressure is proportional to the energy-density, and also that radiometer action can be completely eliminated with a metal vane.

copper and nickel. Tests using high magnetic field-strengths showed that the magnetic properties of this material were no more pronounced than those of such substances as glass and copper. Further, after the erection of the apparatus had been completed the periodic time of the suspension was found to be unaffected by a strong magnetic field. It was therefore justifiable to assume that the deflections of the vane would be unaffected by a slight variation in the magnetic field of the laboratory, if such occurred.

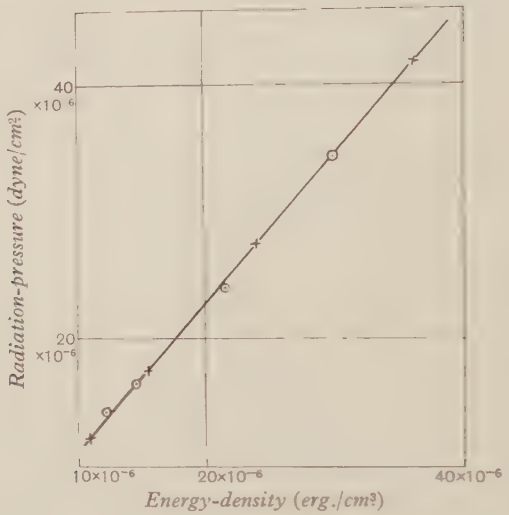


Figure. 12. Results obtained with aluminium vane.

Observations (1)*: evaluation of radiation pressure.* Data relating to the suspension:

Linear dimensions of the vane: length 21·93 mm., depth 11·02 mm.
Moment of inertia of suspension, 0·0903 gm.-cm.²
Periodic time of suspension, 67·6 sec.
Scale distance, 104·3 cm.

Determinations were made of the deflections produced with the radiation incident upon all four end faces of the vane in turn. In this experiment there was found to be a slight regular drift of zero having a period of about 40 minutes. This was very persistent, lasting over several days, and probably arose from pendulum motions of the suspension. By taking the deflectional readings during the intervals of maximum displacement of the zero, i.e. when the rate of change of zero was a minimum, the effect of the periodic wandering was practically eliminated.

In the case of the majority of the observations given in table 13 the beam was quickly turned from one end of the vane to the other without an intermediate checking of the zero. The double deflection was thus directly observed.

Table 13. Deflectional data (mm.)

Beam	Deflections when beam was directed on		Final mean
	left-hand surface of vane	right-hand surface of vane	
1	44·2	46·9	
	45·5	45·7	
	44·8	45·6	
	44·3	45·7	
	44·6	—	
	44·1	—	
Mean	44·4	46·0	45·2
2	64·7	64·8	
	64·6	66·0	
	65·1	63·4	
	62·4	65·0	
	64·8	66·7	
	66·2	65·8	
Mean	64·6	65·3	65·0
3	84·4	91·0	
	85·6	85·6	
	82·7	—	
Mean	84·2	88·4	86·3

There is a rather large discrepancy in the results for the last series. This is probably due to the fact that with a beam of such a high intensity the accurate setting of the spot presented some difficulty.

Reflecting power of the vane. Two of the reasons for using this material were that it took a high polish and that it possessed considerable rigidity in thin layers, so that when once flattened the vane surfaces were not easily impaired. Moreover the material reflected light regularly with very little scattering, and was unaffected by the preliminary heating of the flask. Determinations of the reflection coefficient of the vane, made after the main experiment by means of the radio-balance, gave the results shown in table 14.

Table 14. Reflecting power of the vane (per cent)

Left-hand side		Right-hand side		
Front end	Rear end	Front end	Rear end	Mean
70·7	66·3	62·2	62·8	65·5
65·4	66·6	62·7	64·4	64·8
67·3	64·4	63·0	67·6	65·6
67·7	65·3	61·6	63·9	64·6
69·1	65·5	61·4	63·1	64·8
			Final mean	65·0

Observations (2)*: measurement of the intensity of the radiation.* These determinations were carried out on the completion of the deflectional observations in the manner which obviated the necessity for a separate evaluation of the energy loss due to the flask. In dismantling the latter the two portions of the flask surface, one on each side which had been traversed by the radiation, were cut out. These pieces were in turn placed in front of the radio-balance, at a distance approximately equal to the radius of the flask, while the intensity of the radiation was being measured.

Table 15. Intensity of radiation: radio-balance readings (μW. per 0·03467* cm^2)

Beam	Portion of flask interposed		Mean
	Left-hand side	Right-hand side	
1	2128	2099	2118
	2113	2126	
	2137	2104	
2	3135	3130	3139
	3150	3164	
	3125	3128	
3	4375	4308	4332
	4374	4273	

* Aperture of radio-balance cups.

Some difficulty was generally experienced in setting the vane at right angles to the axis of the optical system. When the fibre was delicate the ultimate position of equilibrium of the suspension could be ascertained only after evacuation of the flask. Moreover a slight displacement was apt to occur when the flask was fused on to the exhausting-apparatus. Sometimes this could be corrected by letting in air, softening one of the connecting tubes, and turning the flask. In this particular experiment the vane was found to set ultimately in a plane deviating by about 15 from the required position. The obliquity of the light converts the relation (12), giving the radiation pressure, into the form

$$(1 + m)\, PAr' \cos \phi = c\theta,$$

l, b in which if l and b represent the length and depth respectively of the vane, the
r' arm r' at which the resultant force operates is given by

$$\tfrac{1}{2}(l - b \sec \phi),$$

ϕ where ϕ is the angle of incidence of the radiation. The cross-sectional area of the beam is calculated as before from relation (9). The value of $\cos \phi$ derived from the directions of the optic axis of the system and the normal to the concave mirror was found to be 0·962. The total correction involved raised the pressure by 8 per cent.† It should be pointed out that this value is a little ambiguous in that it would have been modified by any slight deviation from parallelism between the

† These figures correspond to the zero orientation of the vane. The total correction factor for this position agrees with the mean of the factors for the displaced positions, in the case of a given beam, to about one or two parts in a thousand.

plane of the vane and the normal to the concave mirror, the adjustment of which was only done by eye estimation. Any such deviation would not however have affected previous measurements with the other vanes, in which cases the light was incident normally.

The results finally derived by the use of the alloy vane are as in table 16.

Table 16

Beam	Pressure (dyne/cm²)	Energy-density (erg/cm³)	Difference (per cent)
I	$21{\cdot}2 \times 10^{-6}$	$20{\cdot}4 \times 10^{-6}$	+ 4
II	30·5	30·2	+ 1
III	40·6	41·7	− 3

Distribution of energy in the spectrum. The distribution of energy in the spectrum of the radiation falling upon the various vanes is indicated in figure 13, in which the energy is given in arbitrary units. The curve was derived by means of a Hilger constant-deviation infra-red spectrometer provided with a rock-salt prism. A sharply focused image of suitable size was cast upon the slit by means of the optical projection system used for observation of the radiation pressure. To allow for any selective absorption due to the glass of the flask a piece of the flask used

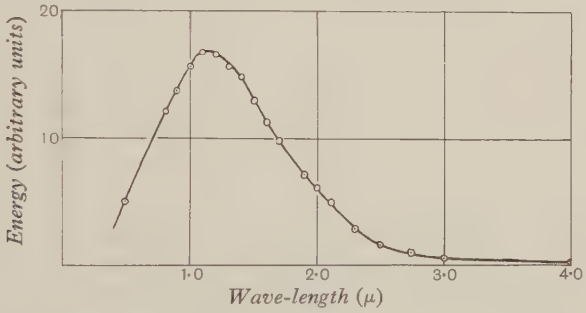


Figure 13. Distribution of energy in the spectrum of the radiation employed.

in the case of the alloy vane was inserted in the path of the beam in front of the spectrometer slit. The filament-brilliancy was the same as in the pressure experiments with the alloy vane. Since the lamp voltage adopted in the case of each of the vanes used was in the vicinity of 100, the curve given may be taken as representative of the quality of the radiation obtaining throughout the whole series of measurements. It will be seen that the wave-length for which the energy is a maximum is in the neighbourhood of $1{\cdot}2\,\mu$, while the bulk of the incident energy arises from the wave band 0·5 to $2{\cdot}2\,\mu$.

The method of varying the intensity of the radiation as required by altering the position of the lamp relatively to the rest of the optical system while maintaining the filament at constant brilliancy preserved the distribution of energy in the spectrum during the pressure and energy-measurements with any given sus-

pension. Similarly the determination of the reflecting power of the vanes with the lamp voltages at the values obtaining during the deflectional readings avoided errors which might arise from the possible variation of reflection coefficients with the quality of the radiation.

Comparison of present experiments with those of previous observers. In comparing the experiments described with those conducted by earlier investigators, we must first direct attention to the work carried out by Nichols and Hull, whose results have hitherto been regarded as conclusive. The methods of measurement which they followed were in general considerably more complicated than those adopted in the present experiments. For example their calorimetric method of determining the radiation-intensity by measuring the rise of temperature produced in a metal disc entailed experiments which were indirect and lengthy in comparison with the use of a Callendar radio-balance for the same purpose. A further complication arose from the fact that Nichols and Hull's source of light was an arc lamp, difficult to keep steady for any length of time. This involved the introduction of an auxiliary bolometer by means of which the displacements of the suspension could be reduced to those which would be given by a lamp of standard intensity. Incidentally this correction, which in some cases was of the order of 50 per cent, involved the assumption that the radiation pressure was directly proportional to the intensity of the beam.

Since the radiation was directed upon the vane for only six seconds, it was necessary to reduce both the apparent intensity of the beam, as determined by the bolometer, and the deflection of the vane to the values which would obtain under static conditions. The reduction necessary in the former case was determined experimentally and involved a correction factor of some 50 per cent. The latter correction was wrongly evaluated and is discussed later.

It may be pointed out that Nichols and Hull used an extremely small deflection, namely 16 mm. on a scale distant 105 cm. from the suspension. Angular deflections up to about fifteen times this value have been employed in the present work.

In attempting to establish experimentally a relation between the pressure and energy-density of radiation it is of fundamental importance to derive a series of measurements covering as wide a range of radiation-intensities as possible. It is noteworthy that Nichols and Hull used what amounts practically to one intensity only. They give one final value for each of three different qualities of radiation, but the two extreme intensities involved differed by only some 8 per cent. Experiments we have conducted with the infra-red spectrometer lead us to doubt whether the quality of radiation obtained by Nichols and Hull by passing full radiation through red glass differed very greatly from the quality of the full radiation itself, unless the glass they used had special properties.

It is to be borne in mind that, when Nichols and Hull undertook their investigation, three modern advantages were not available to them, namely (*a*) exhaustion apparatus capable of producing a vacuum high enough to eliminate radiometer action; (*b*) a steady and powerful source of radiation; and (*c*) a quick and reliable instrument for use in connection with the energy-measurements. Nichols and Hull

employed very elaborate and ingenious methods in the attempt to surmount these difficulties, but it is not to be expected that their indirect methods could give results so satisfactory as those obtained with more modern facilities.

Miss Golsen's work has not received much attention, possibly owing to the fact that Nichols and Hull's results have been generally accepted as quite conclusive. She worked only with metal vanes, whereas Nichols and Hull used silvered glass suspensions. Miss Golsen made an investigation of radiometer action with these metal vanes, and her general conclusions are in agreement with those given in this paper. For her radiation-pressure determinations she used balanced suspensions with the working surface on one side only of the stem. A range of radiation-intensities was, however, employed. The methods Miss Golsen adopted for the determinations of the energy of the beams, and for the measurement of the reflection coefficient of the vanes, were less direct than those in which a radiobalance is used. The possibility of errors of the order of ± 6 per cent in the experimental work are admitted.

A comparison of the torsional constants of the fibres used in the investigations mentioned are given in table 17.

Table 17

Observers	Couple required to turn suspension through one radian (dyne-cm. $\times 10^{-3}$)	
Nichols and Hull	7·6	
Miss Golsen	1·1 to 4·0	
M. B.	Glass suspension	1·3
	Mica suspension	1·25
	Aluminium suspension	0·22
M. B. and S. E. G.	Alloy suspension	0·8

§ 4. ACKNOWLEDGMENTS

The authors are much indebted to the late Prof. H. L. Callendar for the interest which he took in this work. They also wish to record their grateful thanks to Mr W. J. Colebrooke and to Mr C. Hellary, of the Physics Workshop of the Imperial College of Science, and to Mr Baker, of Howard, Rawson and Co., for their help in the constructional work.

APPENDIX

Errors of calculation occurring in Nichols and Hull's work

In view of the great importance in modern physical theory of the relation of numerical equality between the pressure and energy-density of radiation, attention is directed to the experimental verification attempted by Nichols and Hull whose work has been widely quoted as the standard investigation in this connection. Stated by these observers as indicating agreement to 1 per cent between the pressure and energy-density, Nichols and Hull's results have hitherto been generally

accepted as experimentally affording conclusive quantitative evidence of the validity of Maxwell's relation. Examination of the methods adopted by these investigators in deducing their results from their observations shows, however, the presence of considerable errors, inasmuch as corresponding values of pressure and energy-density for radiation of given intensity exhibit, when correctly evaluated, a divergence of some 10 per cent.

Working with a gas pressure of 16 mm. of mercury, Nichols and Hull did not directly observe the angle of displacement of their suspended system at which the couple due to the radiation pressure balanced that due to the supporting fibre. Nichols and Hull assumed that it was impossible to obtain a vacuum good enough to eliminate the disturbing influence of the residual gas surrounding their suspension. They therefore determined experimentally the pressure at which this radiometer action was a minimum, and made their observations at this pressure (16 mm.). Since gas action takes time to develop, whereas radiation pressure acts instantaneously, Nichols and Hull assumed that by keeping the radiation on the vane for six seconds (one quarter of the period of oscillation of the suspended system) gas action was eliminated, although at this pressure the gas caused the motion of the suspension to be heavily damped.

The procedure adopted was as follows. The suspension being initially at rest, the radiation was directed on the vane for one quarter of the periodic time of the suspension. On the removal of the beam, the vane continued to move towards a maximum displacement, which was observed. From this observed amplitude of the first swing of the suspension, the angle of displacement at which the couple due to radiation pressure would balance that arising from the torsion of the fibre could, assuming radiometer action to be negligible, be derived in the following way.

Throughout the interval during which the radiation is incident upon the vane the equation of motion of the suspension may be written

$$K\frac{\partial^2\theta}{\partial t^2} + 2\epsilon\frac{\partial\theta}{\partial t} + G\theta = L \qquad \text{.....(1),}$$

K, G where K and G represent respectively the moment of inertia of the suspension and
ϵ the fibre constant, while 2ϵ is the damping factor, the assumption being made that the sole effect of the residual gas is to damp the vibrations. Since the constant L represents the torque due to the radiation, the angular displacement of the suspension, at which the couples arising from the radiation and the torsion of the fibre are equal, is given by L/G. The solution of the above equation of motion is

$$\theta = \frac{L}{G}\left[1 - e^{-(\epsilon/K)t}\sqrt{(1+x^2)}\cos(2\pi t/T - \delta)\right] \qquad \text{.....(2),}$$

x, T where $x = \epsilon T/2\pi K = \tan\delta$ and $T = 2\pi/\sqrt{(G/K - \epsilon^2/K^2)}$,

the values of the constants being determined by the initial conditions that θ and $\partial\theta/\partial t$ both vanish when $t = 0$.

Hence the values of the displacement and the velocity of the suspension ob-

taining on the removal of the radiation, after the lapse of time interval $\frac{1}{4}T$ from the commencement of the application, are given by the relations

$$\theta_{\frac{1}{4}T} = \frac{L}{G}[1 - e^{-\epsilon T/4K}.x] \qquad \ldots\ldots(3),$$

$$\left(\frac{\partial\theta}{\partial t}\right)_{\frac{1}{4}T} = \frac{L}{G}e^{-\epsilon T/4K}\left(\frac{2\pi}{T} + \frac{2\pi}{T}x^2\right) = \frac{L}{G}\cdot\frac{2\pi}{T}.e^{-\epsilon T/4K}.(1 + x^2) \qquad \ldots\ldots(4).$$

On the removal of the radiation pressure the equation of motion of the suspended system becomes

$$K\frac{\partial^2\theta}{\partial t^2} + 2\epsilon\frac{\partial\theta}{\partial t} + G\theta = 0 \qquad \ldots\ldots(5),$$

with the solution $\theta = Ae^{-(\epsilon/K)t}(\cos 2\pi t/T - \alpha)$ $\ldots\ldots(6),$ A

in which the period T has the same value as before (24 sec. in the practical case), while the new zero of time is the instant of removal of the light. If r denotes the r ratio of successive amplitudes of the damped vibrations of the suspension, then

$$\frac{\epsilon T}{2K} = \log\left(\frac{1}{r}\right) = \pi x.$$

The values of displacement and angular velocity given in relations (3) and (4) may be expressed

$$\theta_{\frac{1}{4}T} = \frac{L}{G}\left[1 - \frac{r^{\frac{1}{2}}}{\pi}\log\left(\frac{1}{r}\right)\right] \qquad \ldots\ldots(7),$$

$$\left(\frac{\partial\theta}{\partial t}\right)_{\frac{1}{4}T} = \frac{L}{G}r^{\frac{1}{2}}\frac{2\pi}{T}\left[1 + \frac{1}{\pi^2}\left\{\log\left(\frac{1}{r}\right)\right\}^2\right] \qquad \ldots\ldots(8).$$

These values determine the magnitude of the constants A and α in (6). Putting this relation in the form

$$\theta = e^{-(\epsilon/K)t}[X\cos 2\pi t/T + Y\sin 2\pi t/T] \qquad \ldots\ldots(9),$$

so that

$$\frac{\partial\theta}{\partial t} = e^{-(\epsilon/K)t}\left[-\frac{\epsilon}{K}\left(X\cos\frac{2\pi t}{T} + Y\sin\frac{2\pi t}{T}\right) + \frac{2\pi}{T}\left(Y\cos\frac{2\pi t}{T} - X\sin\frac{2\pi t}{T}\right)\right] \qquad \ldots\ldots(10)$$

and in each of these equations writing $t = 0$, we find from (7) and (8) that

$$X = \frac{L}{G}\left[1 - \frac{r^{\frac{1}{2}}}{\pi}\log\left(\frac{1}{r}\right)\right]$$

$$Y = \frac{L}{G}\left[r^{\frac{1}{2}} + \frac{1}{\pi}\log\left(\frac{1}{r}\right)\right].$$

Hence $$A = \sqrt{(X^2 + Y^2)} = \frac{L}{G}\sqrt{(1 + r)\left[1 + \frac{1}{\pi^2}\left\{\log\left(\frac{1}{r}\right)\right\}^2\right]} \qquad \ldots\ldots(11).$$

On insertion of the experimentally determined value of r, namely 0·783, derived by Nichols and Hull in the case of their suspension

$$A = 1{\cdot}339\, L/G; \text{ and since } X/A = \cos\alpha, \quad \alpha = 0{\cdot}695 \text{ radian.}$$

From relation (10) the displacement θ has its maximum value when $t = 2{\cdot}77$ sec. Hence

$$L/G = \theta_{\max.}/1{\cdot}260.$$

The angle at which the couple due to torsion of the fibre would, in the absence of radiometer forces, balance that due to radiation pressure is thus given as a fraction of the first amplitude of swing of the suspension.

Errors made by Nichols and Hull in (a) the evaluation of the radiation pressure

Adopting the general method of the analysis given, Nichols and Hull took the differential equation (1) as defining the motion of their suspension while radiation was incident upon it. Instead, however, of using the proper solution of this equation to obtain the position of the vane at any instant, they assumed the displacement θ to follow the relation

$$\theta = \frac{L}{G}\left(1 - e^{-(\epsilon/K)t}\cos\frac{2\pi t}{T}\right) \qquad \text{.....(12).}$$

On this assumption the values of displacement and velocity respectively at time $\frac{1}{4}T$, the instant of removal of the radiation, are given by

$$\theta_{\frac{1}{4}T} = L/G$$

and

$$\left(\frac{\partial\theta}{\partial t}\right)_{\frac{1}{4}T} = \frac{L}{G}\cdot\frac{2\pi}{T}e^{-\epsilon K/1T}.$$

Under these circumstances

$$X = \frac{L}{G} \text{ and } Y = \frac{L}{G}\left[r^{\frac{1}{2}} + \frac{1}{\pi}\log\left(\frac{1}{r}\right)\right]$$

and hence

$$A = \sqrt{(X^2 + Y^2)} = \frac{L}{G}\sqrt{\left[1 + r + \frac{2r^{\frac{1}{2}}}{\pi}\log\left(\frac{1}{r}\right) + \frac{1}{\pi^2}\cdot\log\left(\frac{1}{r}\right)^2\right]}.$$

In their paper Nichols and Hull omit the term $\left(\frac{1}{\pi}\log\left(\frac{1}{r}\right)\right)^2$ as being negligibly small and give the value of $\sqrt{\left[1 + r + \frac{2r^{\frac{1}{2}}}{\pi}\log\left(\frac{1}{r}\right)\right]}$ as 1·357. Correctly calculated on the basis $r = 0{\cdot}783$, the figure given by Nichols and Hull, the value is 1·386. It appears that they have erroneously taken $\log(1/r)$ to the base 10 instead of to the base e.

Having assumed the equation of motion of the suspension subsequent to the removal of the radiation to be

$$\theta = 1{\cdot}357\frac{L}{G}e^{-(\epsilon/K)t}\cos\left(\frac{2\pi t}{T} - \alpha\right),$$

Nichols and Hull make a further error in their deduction from this equation of the value of the first maximum amplitude of swing. Instead of deducing this from the time at which the angular velocity of the suspension has its first zero value, they have assumed that

$$\theta_{\text{max.}} = 1{\cdot}357\, L/G,$$

which is clearly incorrect since the heavy damping of the motion is ignored. Thus, while the true relation, giving the angle for which the couples due to the radiation and torsion of the fibre would balance, is

$$L/G = \theta_{\text{max.}}/1{\cdot}260,$$

Nichols and Hull used the equation

$$L/G = \theta_{\text{max.}}/1{\cdot}357\,.$$

Being proportional to L/G, their stated values of the pressure of the radiation are about 7·5 per cent too low.

It is to be noticed that the error would have been about 10 per cent if A had been correctly evaluated. The arithmetical miscalculation diminished to some extent the effects of the errors of analysis, namely, the adoption of an inaccurate equation (12) instead of (2) for the initial motion of the suspension, and the erroneous derivation of $\theta_{\text{max.}}$ in the last stage.

Errors made by Nichols and Hull in (b) the evaluation of the energy-density.

A further inaccuracy occurs in the estimation of the energy-density of the radiation. The method employed by Nichols and Hull to measure the intensity of the radiation involved the measurement of the rise in temperature produced in a metal disc by the incidence of the radiation. For the value of the mechanical equivalent of heat occurring in the calculation reference was made to Griffith's paper*. From this paper the value of J is wrongly quoted as being $4{\cdot}27 \times 10^7$ ergs per gram calorie, whereas the stated value is 427 gram-metres per gram calorie. Nichols and Hull have thus omitted to make the necessary conversion from gravitational to absolute units. The radiation-intensity values being proportional to the value of the mechanical equivalent, taking J to be 4·18 joules per calorie, have the effect of making Nichols and Hull's figures for the energy-density some 2 per cent too high.

Nichols and Hull's results as finally stated showed the values for the radiation pressure to be slightly greater than those for the energy-density by an amount of the order of 1 per cent. Allowance for the errors mentioned above in each case enhances the divergence, since an increase in the pressure is involved in conjunction with a decrease in the energy-density. It is thus seen that Nichols and Hull's results, when correctly calculated, show the pressure of radiation to be in excess of the energy density by some 10 per cent. Hence they cannot be regarded as furnishing conclusive quantitative evidence of the validity of the relationship of equality deducible from theory.

* *Phil. Trans.* A. 184, 496 (1893).

It should be mentioned that in addition to the above there is considerable ambiguity in Nichols and Hull's paper regarding their use of galvanometers. In the course of their experiments they used two galvanometers G_1 and G_2. G_2 was in the auxiliary bolometer circuit which, as mentioned on page 348, was used to correct for variations of the intensity of the incident beam. G_1 was in circuit with the silver disc whose rise in temperature was employed to determine the energy density. The sensitivities of these galvanometers varied from time to time during the experiments, and therefore it was necessary to reduce all deflections of each galvanometer to those corresponding to the standard sensitivity of each instrument concerned. Nichols and Hull have nowhere defined what they mean by the sensitivity of a galvanometer, but in the following illustration* they have used an inverse definition in the case of G_1, i.e. they take the sensitivity to be the current required to produce unit deflection.

Example. The deflection of G_1 was 393·8 scale divisions for a temperature-range of 39·11 degrees when the sensitivity of galvanometer G_1 was 996. Nichols and Hull state that for a temperature range of 1 degree and with sensitivity of G_1 equal to 1000 a deflection of 10·03 divisions would be produced, i.e.

$$\frac{393{\cdot}8 \times 996}{39{\cdot}11 \times 1000} = 10{\cdot}03.$$

Thus they have used an inverse definition of sensitivity.

Nichols and Hull defined a standard light-beam as one, a constant fraction of which, when reflected into the auxiliary bolometer, would produce a deflection of 100 divisions on the scale of G_2, presumably, though they do not expressly state this, at a standard sensitivity of G_2. The example in table 6 on p. 588† shows that in the case of G_2 they have used the direct definition of sensitivity, i.e. the deflection per unit current. Thus when the energy was being measured, a beam which produced a deflection of 92·4 on the scale of G_2 (sensitivity 667) gave an observed rate of change of deflection on the scale of G_1 (sensitivity 996) of 1·307 mm./sec. The deflection which would be produced by the lamp on the scale of G_2 when at standard sensitivity, the direct definition being assumed, is 92·4 × 1000/667 divisions. The rate of change of deflection on the scale of G_1 at standard sensitivity, the inverse definition being assumed, is 1·307 × 996/1000 mm./sec. Hence the rate of change of deflection of G_1 produced by the standard beam would be 1·307 × 996 × 667 × 100/1000 × 1000 × 92·4, or 0·94 mm. sec.

The above expression is given by Nichols and Hull. All the results on table VII, p. 589† are similarly deduced. (A decimal point has evidently been omitted all through column 3, table VII.) Thus their results are only consistent if we suppose that in this set of calculations they have used two different definitions of the term "galvanometer-sensitivity."

Again the pressure values finally obtained on p. 581† are consistent with the direct definition of galvanometer sensitivity in the case of G_2, though the order in

* *Proc. Amer. Acad.* **38**, 587 (1903).

† Nichols and Hull's paper.

which these values are worked out on pp. 580–581* might at first lead the reader to think that these results were obtained by the use of an inverse definition but were miscalculated.

Thus throughout their work Nichols and Hull's results are only consistent with the adoption of an indirect definition of galvanometer sensitivity in the case of G_1 and a direct definition in the case of G_2. The use, without explanation, of the same term with two different meanings, particularly when occurring in the same numerical expression, introduces complexity into a paper already rather difficult to follow.

It is of interest to observe that if Nichols and Hull had intended to use the usual direct definition throughout, their results would not have been very seriously affected, as the stated sensitivity of galvanometer G_1 was almost of standard value. If, on the other hand, they meant to use always the inverse definition of sensitivity, their stated values of pressure and energy-density would in each case be increased by more than 100 per cent.

At the same time, owing to the fact that the sensitivity of G_2 varied somewhat in the two parts of the experiment, the energy-density values would be increased by about 8 per cent more than the pressure values. This would offset to some extent the divergence between the pressure and energy figures consequent upon the errors of calculation mentioned earlier. However, the absence of precise information regarding the galvanometer-sensitivities renders suppositions such as the above no more than speculative.

DISCUSSION

Dr G. D. West. This paper has been of much interest to me. First, it has administered a mild mental shock. The pressure of light is involved in many physical and astronomical calculations, and hitherto one imagined that it was measurable to 1 per cent. To learn that a quantity so important has, up to now, been measured only to 10 per cent is disquieting. Nichols and Hull did their experiments when our modern vacuum technique was unknown. The pressure of light certainly needs remeasuring with modern apparatus, and I am glad that the authors have taken the matter in hand.

Reading through the paper, one forms the impression that the disturbing gas-action effects are as troublesome as ever. How troublesome they can be I found out some years ago when I made approximate measurements of the pressure of light by allowing a beam to fall on a strip of gold leaf in a test tube. The gold leaf is very light and thin and is deflected by an easily observable amount. There is no appreciable temperature-difference between the surfaces, and one might expect only a negligible radiometer effect. Gas-action effects of an entirely different kind appear, however, and these I subjected to a considerable investigation†. I think the authors would do well to study radiometer effects more generally, and apart from the vagaries of a particular instrument.

* Nichols and Hull's paper.

† *Proc. Phys. Soc.* **31, 32, 33.**

The question of the evolution of adsorbed gas might perhaps be given more prominence. Each molecule as it is emitted gives a reactive kick, and this of course disturbs the measurements. I should like to ask the authors if they are entirely satisfied with the explanation, based on reradiation, that they give of the results in table 3?

Mr G. G. Blake (communicated). On page 323 the authors give 10^{-6} mm. as the pressure at which gas action vanished. Would it not be well to make it quite clear that these figures only apply in the case of a radiometric vane the two faces of which are equidistant from the walls of a spherical containing vessel? Dr G. D. West* has shown that if a strip of aluminium foil is suspended near to one side of an evacuated vessel, instead of centrally, it is, when irradiated, deflected away from that side of the vessel independently of the side of the strip upon which the radiation falls.

In June last year whilst conducting some experiments with a "radiometric condenser†" I employed a thin leaf of aluminium foil to act as the variable plate of a condenser. Both this leaf and a rigid back plate in close proximity thereto were enclosed in an evacuated vessel and employed to control the frequencies of a zero-shunt thermionic valve circuit; the movements of the leaf were then magnified and registered with aid of a galvanometer in the plate circuit of the valve. The glass vessel containing the condenser was exhausted for me by the courtesy of the G.E.C. Research Laboratories to 10^{-7} mm. Incident radiations caused very good excursions of the leaf (when thus magnified) away from its back plate towards the source of radiation. Exhaustion was not carried beyond this value, but certainly we had not nearly reached the limit for this particular radiometric arrangement.

The illustrations of the apparatus employed by the authors do not show any heat-filters between the source of radiation and the radiometer. Is it not possible that the reluctance of the vane to return to zero, mentioned on page 332, may be due (at any rate partially) to increased temperature of the wall of the radiometer flask, on its side nearest to the incident radiation? A repetition of a few of the observations with the radiometer flask protected from the heated atmosphere by means of a Dewar flask might yield different results. In any case the protection of the walls from changes in atmospheric temperature would appear to be advisable. On page 333 it is interesting to compare the figures given by the authors with those given by G. D. West for maximum thermokinetic effects. The authors give 10^{-2} to 10^{-3} mm., and West gives 2×10^{-2} mm. for his radiometric aluminium leaf.

Authors' reply. It is to be strongly emphasized that the observed gas effects, which were perfectly definite and reproducible, apply only to flat rectangular vanes suspended by means of a quartz fibre, centrally, within a spherical glass vessel. The experiments were repeated several times with flasks of different diameters, and the results were always as indicated. The dimensions of the flask were, however, purposely always made great compared with those of the vane, in order to minimize any possible effect due to the proximity of vane and walls. The experiments of

* *Proc. Phys. Soc.* **28** (1916). † *Elec. Rev.* **112**, Jan. 13 (1933).

Mr Blake and Dr West appear to have been widely different: for instance Dr West used gold leaf in a test tube. Divergences in observed effects were therefore to be expected.

In reply to Dr West: The impression that disturbing gas effects were as troublesome as ever is erroneous. One of the purposes of the present paper is to show that with thin metal vanes, at a pressure of 10^{-6} mm., radiometer action was entirely eliminated.

In tables 6 and 11 it is seen that, with these vanes, the conditions of vibration conform exactly to those stipulated in the last two paragraphs on page 323 of the paper. We consider that we have here clear experimental evidence that radiometer action is absent. This evidence is rendered complete when the results given in figure 12 and discussed on page 343 are considered. The question of adsorbed gas is dealt with on page 334. In the course of a long series of observations we recollect only one occasion when a sudden jerk in the deflection implied the emission of gas from the surface of the vane. With regard to the point about reradiation, as is indicated in the paper (page 330) this is merely a suggestion, but reradiation appears to operate in the right direction. We hope to undertake some more work in this connexion, as further experimental data are required.

The first part of Mr Blake's criticism has already been dealt with. It should have been mentioned in the paper that, throughout the observations, the flask was protected from draughts and extraneous radiation by a thick asbestos jacket with suitable openings for admitting the radiation and studying the deflections. It is not easy to see how the slightly increased temperature of the wall of the flask, adjacent to the source of radiation, could produce the regular zero-lags observed. These lag effects and also the initial variation of the deflections with glass vanes (tables 2 and 4) can be explained if it is assumed that the pressure due to true radiometer action (page 333) is gradually counterbalanced by a lowering of the density on the hotter side of the vane; and that similarly, when the radiation is removed, the radiometer action dies away before the density again becomes uniform. If it is remembered that the blackened or platinized film is always the hotter side of the vane, the initial variation of the deflections (tables 2 and 4) when the radiation is first applied, and the direction of the residual deflections (zero lag) when the radiation is removed (table 5), can be readily understood.

REVIEWS OF BOOKS

Faraday's Diary: being the Various Philosophical Notes of Experimental Investigation made by Michael Faraday, D.C.L., F.R.S., during the years 1820–1862. Vol. 1 (Sept., 1820–June 11, 1832), pp. xxiii + 430; vol. 2 (Aug. 25, 1832–Feb. 29, 1836), pp. xvii + 467. (London: G. Bell and Sons. To be completed in seven volumes.) Price twelve guineas the set.

The *Diary*, which was bequeathed by Faraday to the Royal Institution, is published under the editorial supervision of Mr T. Martin. The Foreword is by Sir William Bragg, who has already published—some months ago—an admirable summary of the chief features of the *Diary*.

There are diaries of many kinds, and it would be difficult to imagine anything sharper than the contrast between this of Faraday and that—also posthumously published—of one of his remotest predecessors in the Fellowship of the Royal Society. Faraday's *Diary* is above all else a record of his day-to-day work in the laboratory of the Royal Institution and—what is vastly more important—of the ideas which were continuously inspiring him to this work. Read in conjunction with the perfected accounts in the volumes of *Experimental Researches*, it tells us of Faraday as a scientist everything that can well be conveyed by the printed word. But of Faraday the man, his personal affairs and his personal relations, it tells practically nothing at all.

We do find, it is true (para. 1503) a reference to a visit by Daniell to the laboratory, coupled with a cordial appreciation of Daniell's generosity in a matter of scientific priority —but personal references of this or any sort are extremely rare. This is not, however, to say that the *Diary* lacks human interest. It would be difficult for anyone, however slight his familiarity with the normal routine of life in the laboratory, to read these pages without forming a vivid picture of Faraday at work, of his attitude in the face of difficulties and his reactions to a success or a temporary check. The *Diary* is illuminating in minor details of Faraday's work as well as in the main lines of his researches; take, for example, this almost trivial detail (para. 1941) with reference to the design of a new voltaic battery which was to replace the one already in use: "Must have a plug at the corner of trough to withdraw the acid by." It is by such notes as this that we realize some of the difficulties under which experimental work was prosecuted a century ago—namely, how many types of apparatus, now accepted as commonplace furnishings of even the worst-equipped laboratory, had then to be specially designed for the work in hand. When it is remembered in addition—and this point is not brought out in the *Diary*—how much of Faraday's time was absorbed in routine duties and in Government work on optical glasses, it becomes even more difficult to express in finite terms a proper appreciation of his intense scientific activity during the period covered by these two volumes (1820–1836).

The *Diary* will necessarily and strongly appeal to all—and not only to chemists and physicists—who would like to know more of the processes by which a great experimental philosopher arrived at the results which have had such far-reaching importance, not only in the history of scientific thought, but in the whole of our social and industrial life. These volumes should do much to stimulate a greater interest in the history of scientific discovery—a service particularly necessary at a time like the present, when, in Physics at least, so much of the older work, in spite of its beauty and elegance, is sadly out of fashion.

Mr Martin's editing has been both pious and scholarly. There has been a minimum of interference with Faraday's wording and punctuation, and the placing of the diagrams

in relation to the text follows the original as far as is practicable on the printed page. The cross references which have been inserted (from the *Diary* to the *Experimental Researches*) are a particularly useful piece of editorial work.

The reproductions of Faraday's line drawings, which form an essential part of the *Diary*, deserve a word to themselves. It is conceivable that a professional art critic could trace in the drawings some indication of the teachings of J. J. Masquerier, but to the untrained eye—if this may be said, with all due reverence—they do not differ appreciably in conception or execution from the pictures which most of us draw on blackboards, in rough note-books, and on the blotting-paper provided at meetings. They are all the more delightful on this account, for at least in this one insignificant respect they bring Faraday into closer fellowship with the humblest of his followers. They have been slightly reduced in size for printing, and the reproduction and placing have been performed with conspicuous skill and taste.

Faraday's services to the Royal Institution are matters of old and well-known history. The many demands for his professional services gave him opportunities for acquiring wealth on a scale which might have tempted even men who were far less familiar than he was with poverty. It was not solely his passion for unfettered research which determined his refusal of these demands—it is certain that his conception of his duty to the Institution which sheltered his researches was a very important factor. It is proper to recall here also his possibly over-sensitive pride in the special position of the scientific pioneer—a feeling which made him averse to such forms of public or official recognition as were the common rewards of success or eminence in other spheres.

It follows therefore that perhaps the most fitting of all memorials to Faraday and to the survival of his ideals is to be found in the continued activity of the Royal Institution and its Davy-Faraday Laboratory in the winning and dissemination of new knowledge. In ordering the publication of his *Diary*, the Managers of the Royal Institution have paid to his memory a very appropriate and a more definitely personal tribute. If, in so doing, they have liquidated some part of the Institution's debt to Faraday, they have at the same time done a service to science and scientific history which we are glad to acknowledge here.

H. R. ROBINSON

Wave Mechanics: Elementary Theory, by J. FRENKEL. Pp. viii + 278. (Oxford: Clarendon Press.) 20*s*.

This volume, although intended to be the first of three, is complete in itself. It takes the form of a survey of the subject, leaving advanced mathematical methods for the second volume, and details of the applications to the third.

Its scope, then, is practically identical with that of a number of one-volume accounts of wave mechanics recently published, in that they all consist of an introduction, but only an introduction, to the whole subject. It is in the early part of these books, the justification for assuming a wave-corpuscle parallelism both for light and for matter, that the main divergencies in treatment occur as between different authors.

Frenkel's outlook has returned more closely to the original one of de Broglie than is now usual, although he goes even further in some ways than de Broglie himself. After an illuminating investigation of Newton's corpuscular theory of light, in which appeal is made in a distinctly un-Newtonian manner to the principle of conservation of energy, he proceeds to sketch the underlying ideas of the modern relativity theory, and from this develops Einstein's theory of light darts.

The argument for wave mechanics, summarized, is based on the relativity relations (1) momentum g of a particle $= mv$ and (2) its energy $E = mc^2$, where m is the mass and $= m_0/(1 - v^2/c^2)^{\frac{1}{2}}$, v is the velocity of the particle, and c is the velocity of light. Now if

$v = c$, and the mass is not infinite, then m_0 must vanish, so that particles of zero rest-mass could only exist when moving with the speed of light. Consequently, if they exist at all, they must be corpuscles of light, as is indeed suggested by certain properties of light. For them, the general relation $g = Ev/c^2$ reduces to $g = E/c$. Now from interference and other experiments we are convinced that a wave-length λ is associated with any beam of light, and we are in the habit of defining a frequency ν by the relation $c = \nu\lambda$. If, now, the energy E depends only on the frequency or wave-length, then the relativity condition of covariance shows that the relation between them must be $E = h\nu$.

We may profitably examine this argument in closer detail. E is the energy of an individual corpuscle—photon, let us say—and it cannot depend either on the velocity or on the rest-mass, since each of these is the same for all photons. Consequently it can only be a function of λ, or (what is essentially the same thing) of ν. This is the statement in the book under review, and it has the merit of appearing rigorous, and of showing that the relation $E = h\nu$ is inevitable on logical grounds, without any need for an appeal to experiment. It is easy to see, however, that there is a lacuna which, according to taste, may be either supplied by the reader or regarded by him as one loophole which will remain open for the construction of yet a new theory when, if ever, wave mechanics comes to need a successor. The possibility which is overlooked is that the energy per corpuscle might depend on the number present. This would be the case, for example, if photons were able to influence each other gravitationally or magnetically.

To return to Frenkel's argument. It is pointed out that the equations $E = h\nu$ and its companion $g = h/\lambda$ remain formally true even for particles with a finite rest-mass if a quantity λ and a quantity ν are associated with them, so that it is justifiable to take the step of associating a wave, not indeed with a single particle, but, as in the case of light, with a whole stream of identical particles all having the same momentum and the same energy. If it is assumed that the square of the amplitude of the wave so introduced is proportional to the particle density, then the only step which remains is to assign a physical reality to the wave which already has full mathematical justification. The relations previously quoted are sufficient to find the wave velocity, and to show that there is dispersion and hence a group velocity, which is identified with the particle velocity.

This part of the book has been treated in considerable detail, because it puts the matter in a neater form than most accounts have been able to do. It is, in fact, concerned to show that the fundamental ideas of wave-mechanics are inevitable, whereas writers have generally been concerned (or at any rate able) to make them plausible.

It is to be noted, moreover, that the train of reasoning used imposes a principle of indeterminism, in the sense that the exact position of discrete particles cannot be deduced if only their mean number per unit volume is specified by a continuously distributed quantity, the wave amplitude.

After the development of these ideas, the experiments which have shown the "wave" aspect of electrons and of atoms are briefly described, but the outlook is not that of the experimentalist and but little weight seems to be attached to these mere verifications of a theory which has been established on quite other grounds, and really requires no verification.

As has been mentioned above, the principle of indeterminism is based here, in the first place, on the necessity of allowing a simple harmonic wave to represent a whole stream of similar particles—or rather, copies of one particle, since there is no interaction between them. When the motion of a single electron is to be studied, it is natural to replace the average number per unit volume by the probability of finding the single one in the volume element concerned.

Incidentally, the facts of partial reflection and partial transmission of a beam of a light at an interface between two media, requiring some photons to move one way and some another, are likewise well explained by a probability interpretation. The analogy between

this behaviour and the "fits of easy reflection" and "fits of easy refraction" to which Newton's corpuscles were subject is striking and is well brought out by Frenkel.

The point of view from which Heisenberg approaches the uncertainty principle is set forth in detail, but the author is unsympathetic to it, considering that the part played by the experimenter is stressed too much. He maintains that since all measurements must alter the quantity observed (e.g. in the case of temperature-measurement, conduction by the thermometer or thermocouple enters into the question), therefore the imaginary γ-ray and other measurements cited by Heisenberg do not undermine our faith in determinism by the mere fact that they alter the quantities to be measured. In his view, Heisenberg's argument merely brings out the indeterminacy already fundamentally accepted in the course of the discussion. The difference between these and "classical" experiments is that in the latter the alterations caused by the observation itself can be corrected for; in wave-mechanics they cannot, and this is because of the underlying lack of determination. In his own words "the application of this principle to the phenomena of the motion of elementary particles does not lead to indeterminism . . . but *simply reveals the indeterminism which is pre-existing in these phenomena* . . ."

It is interesting to notice that one of the arguments which has been much used by some writers as a reason for our belief in indeterminacy is hardly even mentioned by Frenkel, viz. the argument that since a wave-packet of finite dimensions must (by ordinary mathematical reasoning) spread as time goes on, therefore the position of the particle which it represents becomes more and more indefinite with the passing of time.

The following section introduces the quantum conditions, illustrated by the motion of an electron in a closed space, so that its associated wave is a stationary one with nodes at the boundaries of the space. This leads up to the formulation of Schrödinger's wave equation, with a further dissertation—an interesting one—on the probability interpretation.

Roughly half the book is devoted to the problems of the single particle in various fields of force, together with the discussion of the philosophical basis of the whole theory. This first half contains, in addition to what has already been discussed, a development, mainly on orthodox mathematical lines, of the theory and its applications to the hydrogen-like atom, the linear and spacial oscillators, the rotator, a very simplified type of crystal lattice and other problems. The "leakage" of particles past a barrier or valley of potential, first discovered in connexion with radioactivity, is applied also to explain the Ramsauer effect and the extraction of electrons from a cold metal by an electric field. Incidentally, the method of replacing the graph of a complicated potential function by an approximation consisting of straight lines is dignified by the name of the Potential Staircase Method.

An important matter is the case where two frequencies are present in the wave function of an electron in an atom. As is well known, these "beat" with the frequency of the light emitted in a Bohr transition from the state associated with one of them to the state associated with the other. In 1924 or 1925, Schrödinger believed that this indicated a fluctuation of a physical charge-density, leading by classical electrodynamics to the emission of light with the frequency of the fluctuations. It is pointed out that this conception cannot be incorporated into wave mechanics with a probability interpretation, because the fluctuations are then definitely debarred from being pictured as movements in a physically real charge-distribution. The author reaches the conclusion that "the important point is not to mix corpuscular conceptions with wave conceptions. . . . If matter is to be described in corpuscular terms . . . then light must also be described in corpuscular terms, and light-emission visualized as the emission of a photon. If, on the other hand, we wish to describe light-emission as the spreading of electromagnetic waves, then the electron must likewise be described as a wave system." Actually the author makes no secret of the fact that he regards the wave picture as definitely superior, both for light and for matter.

No space has been left to discuss the second half of the book, which deals first with the

wave mechanics of a system of particles, and then, proceeding via Pauli's exclusion principle and the properties of symmetric and anti-symmetric functions, with that form of statistical mechanics which is found to be appropriate to wave mechanics. It will be found interesting to read a very lucid paper by Brillouin* in connexion with this chapter, since they are in some respects mutually complementary.

The last chapter differs from the rest of the book not only in subject-matter but in style. It was written after the proofs of the rest were completed, and is in a sense a foretaste of the third volume. It deals with the applications of quantum statistics, chiefly to the properties of metals, and contains a certain amount of original matter but no answer to Fowler's question† "Why is a metal?"

J. H. A.

Short Wave Wireless Communication, by A. W. Ladner, A.M.Inst.C.E., and C. R. Stoner, B.Sc., A.M.I.E.E. Pp. xii + 348, with 13 plates and 201 text figures. (London: Chapman and Hall, 1932.) 15*s*.

This is a very useful general review of the methods and problems of radio communications using wave-lengths less than 150 metres. That it is mainly concerned with the Marconi systems is natural in view of the experience of the authors and, although this does not lead to the exclusion of some reference to other systems, the fact might in the next edition be recognized in some such sub-title as "with special reference to the Marconi systems." It has all the merits and some of the defects of a text-book which has grown from lecture notes; the diagrams are blackboard diagrams, and the authors sometimes forget that they are no longer beside the blackboard to lighten its darkness. Figures 54, 58, 60, 85 and 120 are typical of the cases in which exposition is inadequate or misleading, labels of ordinates and scale values are missing, and so on. The authors have learned that repetition is a virtue in teaching, and have not been afraid of it, but it is doubtful whether the degree of colloquialism and anthropomorphism which is permissible in the lecture room is appropriate to the printed page. It is, for example, hardly fair to the reader to say (p. 94): "Actually no intelligence can be conveyed by a single side-band system unless (*a*) The receiver knows what intelligence it is to expect...." There is, too, a certain carelessness about proper names ("Delinger," p. 67, and "Vigoreaux," p. 173, for example), about units (microvolts per metre are not the appropriate units for internal noise-levels, kilocycles are not units of frequency), and about good English ("flick impulsing," p. 118, is an unhappy addition to technical vocabulary, and "parasites" *tout court*, p. 146, are irritating...). As a last item in this critical paragraph we would protest against this statement: "The name 'push-pull' is given to those circuits impulsed by two valves arranged differentially. Of course it is not necessary to employ two valves to obtain a push-pull action...." This may be true, but in this shape it is not good for students.

But when all this has been said we must repeat that we like the book, and we may cite the chapters on high frequency feeders, aerial arrays, commercial receivers, commercial transmitters, and ultra short-waves, and the paragraphs on Eckersley-Tremellen Charts as especially interesting sources of data not conveniently available elsewhere.

Optische Abbildung: Einführung in die Wellen- und Beugungstheorie optischer Systeme, von Dr Johannes Picht. Pp. x + 243, with 65 figures. (F. Vieweg und Sohn, Brunswick, 1931.) Paper covers, RM. 20; bound, RM. 22.40.

Dr Picht has been engaged in recent years in carrying out an extensive detailed investigation of the properties of images formed by optical systems according to the wave

* *Annales de Physique*, 7, 315 (1927). † *Proc. Phys. Soc.* 42, 458 (1930).

theory of light. Although the physical principles to be employed are simple, the examination of the light-distribution in the image space in the presence of aberrations inevitably involves somewhat elaborate mathematical expressions. Those who are interested in the subject should certainly consult this book, which summarizes much of the work already done. The fact that some numerical results are presented in graphical form will be particularly appreciated by those who do not care to follow the more general investigations. We miss references to somewhat similar work published in this country.

Grundlagen der praktischen Optik: Analyse und Synthese, optische Systeme, von Dr M. Berek. Pp. vii + 152, with 63 figures including one plate. (Walter de Gruyter und Co., Berlin, 1930.) RM. 13.

A large part of this book is devoted to the exposition of general optical laws and of methods of calculation such as form an important feature of a number of German optical books. It differs from some others in presenting and applying the convenient thin-lens formulae obtained by Airy and others which became widely known through such books as Coddington's and Dennis Taylor's. In addition a system of equations is given for the direct solution of triple photographic lenses, which is claimed to be simpler than the systems of some other German writers. These equations only cover the first-order aberrations of thin lenses, and no guidance is given for systems of more than three lenses. Numerical examples are included, but the discussions throughout are confined to well known systems. The book is reasonably indexed and a table of symbols and their meanings is included.

Atomic Energy States, by R. F. Bacher, Ph.D., and S. Goudsmit, Ph.D. Pp. xiv + 562. (New York and London: McGraw-Hill Publishing Co., Ltd.) 36*s*. net.

Since the appearance of A. Fowler's *Report on Series in Line Spectra* and Paschen and Götze's *Seriengesetze der Linienspektren* in 1922, great progress has been made in the theory and analysis of atomic line spectra, and several important books on certain aspects of the subject have been published, such as Hund's *Linienspektren und Periodisches System der Elemente* (an English translation of which should certainly have been published in 1926 or 1927), Grotrian's *Graphische Darstellung der Spektren* in two volumes, Pauling and Goudsmit's *The Structure of Line Spectra*, and Ruark and Urey's *Atoms, Molecules and Quanta*. The last two of these belong to the International Series to which has recently been added the volume now reviewed. An adequate idea of the enormous output of papers on atomic spectra in this period can only be obtained by perusal of the extensive bibliography accompanying Gibbs's review, *Line Spectra of the Elements*, the first part of which appeared last year in *Reviews of Modern Physics*.

Of the works mentioned, none but Bacher and Goudsmit's supplements the 1922 tables of Fowler and Paschen-Götze with a new compilation of complete numerical data. In the present work we find extensive tables giving the classifications (symbols, electron configurations, etc.) and numerical values of the known terms for all the atomic emitters, neutral and ionized, arranged in alphabetical order, from neutral A to trebly-ionized Zr. These tables, with brief notes at the head of each, occupy no fewer than 519 of the 562 pages; many pages, it is true, are less than half-filled, thus allowing useful space for the reader to add new data as they appear in the journals. In an introduction of 21 pages an adequate account is given of such results of atomic spectrum theory (e.g. electron configurations, couplings, selection rules, etc.) as are necessary for the understanding of the notation and the construction of the tables. The tables, unlike those of Fowler and Paschen-Götze, do not include wave numbers of observed lines; to obtain an idea of the

spectrum of a given emitter from the relevant table of term values, the reader must apply results summarized in the introduction. A very welcome and useful inclusion near the end of the book is Paschen's 1928 tables of Rydberg terms $xR_\infty/(m+a)^2$, with x equal to 1, 4, 9 and 16.

The compilers have wisely avoided mention of any notation but the one now in general use for atomic spectra. This is in pleasing contrast to the rather unnecessary inclusion in some other recent spectroscopic works of detailed discussions of the two notations and empirical systems of numeration which served adequately ten years ago but have now given place to those used in this volume.

The printing work is excellent throughout. It may be pointed out, however, that 1_2, 1_2^1, 2_2^1 appear on page 6 for $\frac{1}{2}$, $1\frac{1}{2}$, $2\frac{1}{2}$. Again, English readers are hardly accustomed to the placing of the decimal point at a lower, rather than a higher, level than a dot used as a sign of multiplication; thus the familiar form "$h = 6{\cdot}547 \,.\, 10^{-27}$" or "$6{\cdot}547 \times 10^{-27}$" appears as "$h = 6.547 \cdot 10^{-27}$," and so on throughout the book.

The tables are very well arranged, and the introductory chapter is clear and logical apart from the fact that the expressions "equivalent electrons," "dipole radiation" and "quadrupole radiation" are used without explanation or definition. Many of the sentences are unnecessarily involved and inelegant on account of the excessive use of the word "one" (apart, of course, from its necessary occurrence as a numeral) as in the following passages:

p. 5: "one places the electrons in order of their n, and for the same n in order of l. For many electrons one usually gives the symbols for only the outermost ones. . . . When the arithmetical sum of all the l's of the electrons is *even*, one obtains *even* energy levels; and, in the other case, *odd* ones."

p. 11: " . . . if one considers two electrons, one in a low energy state and the other in a highly excited state, the spin-orbit interaction of the first may be larger than its interaction with the second one. If one now brings the second electron. . . . In such a case low states will show Russell-Saunders coupling; higher ones will not."

Similarly, on pages 3 and 6 "each one of" and "the ones" are used for "each of" and "those" respectively.

These minor blemishes, however, will in no measure detract from the very great value of the work to the active spectroscopist; to him, indeed, it is well-nigh indispensable. It will be used to best advantage in connexion with other works, such as Grotrian's second volume, Pauling and Goudsmit's book and Gibbs's bibliography.

W. J.

Annual Tables of Constants (A.T.C. and Numerical Data). (Paris: Gauthier-Villars et Cie.) Vol. 8, part 2, 1927–1928. Pp. xxi + 1604. 500 fr. Parts 1 and 2. Vol. 9, 1929. Pp. l + 1607. 400 fr. Index to vol. 9.

Physicists need no introduction to the volumes which mark the progress of Dr Marie's great enterprise. With the increase in output of scientific research has come a corresponding increase in the dimensions of the *Annual Volumes*, and now the data of vol. 8, part 2 (for 1927–1928) and vol. 9 (for 1929) are contained in two massive volumes which together turn the scale at nearly 18 pounds' weight. But the editor, realizing that a guide to this labyrinth is very necessary, has provided an admirably arranged *table des matières* to vol. 9, which considerably reduces the time and trouble of running down a particular figure.

It is impossible, and undesirable, to give a detailed review of the contents of these weighty tomes. It is sufficient to say that the reviewer has tested this immense mass of chemical, physical and biological data at various points, and has nowhere found the

desired information wanting, or over-difficult of access. The tables are indispensable to all serious researchers, and it is by no means one of their least merits that the publication is produced at a price which brings them within the reach of individual workers.

A. F.

Photograms of the Year 1932, edited by F. J. MORTIMER, Hon.F.R.P.S. (London: Iliffe and Sons, Ltd.) Paper covers, 5*s.*; cloth bound, 7*s.* 6*d.*; half-bound leather, 10*s.* 6*d.*

Photograms of the Year provides a very useful standard by which the amateur in photographic art may learn to assess his work. A series of more than eighty beautifully reproduced photographs affords ample practice for the study of grouping, of balance, of chiaroscuro and of choice of subject.

The value of the book is enhanced by an article from Mr C. J. Symes commenting critically on the pictures reproduced in the volume, and by a summary of the year's work written by the editor.

A. F.

DR W. JEVONS' REPORT ON

BAND-SPECTRA OF DIATOMIC MOLECULES

308 pp. Numerous illustrations, spectrograms and tables of numerical data

Paper covers 17*s.* 6*d.*; *post free* 18*s.* 0*d.*

Bound in cloth 20*s.* 6*d.*; *post free* 21*s.* 0*d.*

PUBLISHED BY

THE PHYSICAL SOCIETY

1 LOWTHER GARDENS, EXHIBITION ROAD

LONDON, S.W.7

The REVIEW of SCIENTIFIC INSTRUMENTS

with PHYSICS NEWS AND VIEWS

F. K. RICHTMYER, *Editor* (*Cornell University, Ithaca, N.Y.*)

PUBLISHED MONTHLY by the American Institute of Physics in collaboration with the Optical Society of America and the Association of Scientific Apparatus Makers of America.

INCLUDES, beginning January, 1933, Editorials and special articles on recent developments in physics; also news of interest to physicists; tables of contents of current journals.

FEBRUARY, 1933 — CONTENTS — VOL. 4, NO. 2

Physics Forum:
- The American Institute of Physics.
- The Statistical Behavior of Photons.
- The Horse-Show Aurora.
- Vibrational Isotope Effects in Polyatomic Molecules.
- The Velocity of Sound in a Gas.

Contributed Articles:
- On the Quantitative Theory of the Wimshurst Static Machine—I.
- On the Quantitative Theory of the Toepler-Holtz Static Machine—II.
- Notes on the Weston Photronic Photoelectric Cell.

Contributed Articles: (*cont.*)
- Eclipse Cinematography.
- Rotary Motion with Friction.

Abstracts of articles dealing with instruments and methods.

News and Views:
- Book Reviews.
- Notes.
- News of the Societies.
- Calendar of meetings and other events of interest to physicists.

Current Literature of Physics

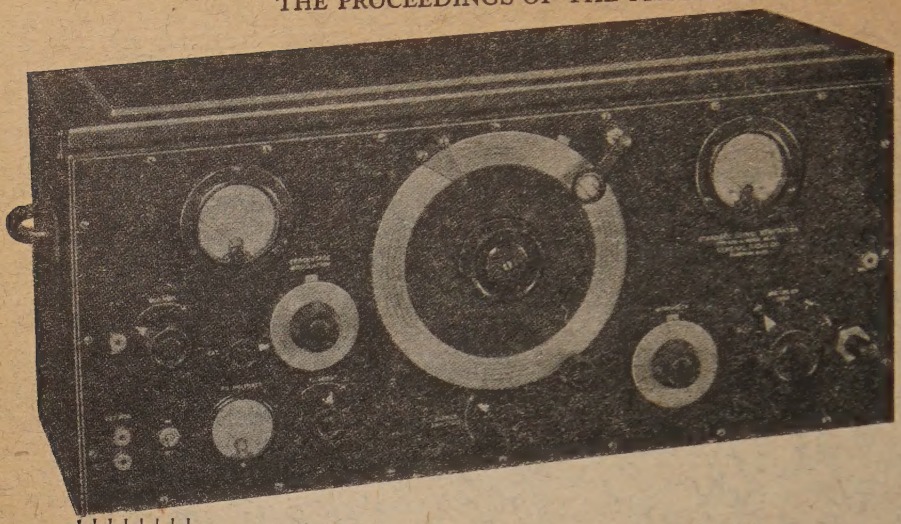
SEND SUBSCRIPTIONS TO

AMERICAN INSTITUTE OF PHYSICS

11 EAST 38TH STREET, NEW YORK, N.Y., U.S.A.

ANNUAL SUBSCRIPTION: Four dollars per year post paid.

COMBINED SUBSCRIPTION: with the Journal of the Optical Society of America, *Six dollars per year.*



As Demonstrated at our Stand No. 49 1933 Exhibition

NEW INSTRUMENTS

Hardly a week goes by—certainly never one single month—without we announce some new and interesting piece of Precision Laboratory Equipment. Here is illustrated, for example, our new "G.R." Type 603-A Standard-Signal Generator, range 12 Metres to 3000 Metres. This is the latest of a World-Famous string of Generators. It is well-nigh perfect—the answer to the Radio-Engineer's fervent prayer.

If Engineers and Physicists wish to keep closely in touch with the latest developments in the Radio, Electrical, Musical Reproduction, Sound Recording, Communications, Broadcasting, Electro-Therapeutical, Telephone Engineering and Allied Fields, then they should keep closely in touch with Claude Lyons, Ltd. We introduce more apparatus each year than all competitive British instrument makers lumped together.

Many of these new Instruments are the absolute pioneers of their type. In time they certainly get more or less copied. For example, for three years our competitors have been struggling to duplicate our first and now very obsolete Signal-Generator, but we are still alone, in England, in this field. *Why not request our New 174-Page Catalogue "G" to-day?*

CLAUDE LYONS LIMITED

Head Offices: 76 OLDHALL ST, LIVERPOOL
London Offices: 40 BUCKINGHAM GATE, S.W. 1

PRINTED IN GREAT BRITAIN BY WALTER LEWIS, M.A., AT THE UNIVERSITY PRESS, CAMBRIDGE